

GenCore version 4.5
Copyright (C) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 2, 2002, 18:02:22 ; Search time 13.08 Seconds
(without alignments)
679,734 Million cell updates/sec

Title: US-09-886-400-4

Perfect score: 1877

Sequence: 1 LRALVFHGNLOYAIEPKSEI.....RRLDAFRAIYNDRGNGEP 364

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 2442594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1580	84.2	346	2	US-08-613-220B-4
2	141.5	7.5	647	1	US-07-894-212A-8
3	141.5	7.5	649	1	US-07-894-212A-2
4	141.5	7.5	650	1	US-07-893-928A-1
5	104	5.5	329	1	US-08-270-013B-2
6	104	5.5	329	1	US-08-838-418-2
7	88.5	4.7	2089	1	US-08-418-893D-23
8	88.5	4.7	2089	1	US-08-418-893D-24
9	87	4.6	235	2	US-09-141-135-2
10	87	4.6	657	4	US-09-355-166-1
11	86.5	4.6	689	1	US-08-248-021A-2
12	85.5	4.6	764	4	US-09-235-451-36
13	85.5	4.6	788	2	US-08-907-166-6
14	83.5	4.4	641	1	US-07-718-535-3
15	83.5	4.4	641	1	US-08-161-999-3
16	83.5	4.4	1676	4	US-08-487-283A-2
17	83	4.4	506	2	US-08-849-480A-5
18	83	4.4	636	4	US-09-564-805-237
19	82.5	4.4	549	2	US-08-676-279-59
20	82	4.4	303	3	US-09-002-298-1
21	82	4.4	623	4	US-09-104-068-4
22	82	4.4	637	4	US-09-104-068-2
23	81.5	4.3	491	1	US-08-489-733-3
24	81.5	4.3	491	2	US-08-993-581B-3
25	81.5	4.3	806	1	US-08-451-715A-6
26	81	4.3	1065	4	US-09-412-545-2
27	80.5	4.3	550	1	US-08-484-493-2

28	80.5	4.3	550	1	US-08-484-494-2	Sequence 2, Appli
29	80.5	4.3	550	2	US-08-345-212-2	Sequence 2, Appli
30	80.5	4.3	550	4	US-09-249-003-2	Sequence 2, Appli
31	80.5	4.3	554	3	US-08-904-871-1	Sequence 1, Appli
32	80	4.3	855	2	US-08-468-558-2	Sequence 2, Appli
33	80	4.3	855	4	US-08-676-444-2	Sequence 2, Appli
34	80	4.3	1110	1	US-08-118-441-29	Sequence 29, Appli
35	80	4.3	1110	3	US-08-338-579A-29	Sequence 29, Appli
36	80	4.3	1110	5	PCT-US94-09851-29	Sequence 29, Appli
37	79.5	4.2	831	1	US-08-073-384C-5	Sequence 5, Appli
38	79.5	4.2	831	1	US-08-254-359A-5	Sequence 5, Appli
39	79.5	4.2	831	1	US-08-483-043-5	Sequence 5, Appli
40	79.5	4.2	831	1	US-08-481-238-5	Sequence 5, Appli
41	79.5	4.2	831	2	US-08-471-066B-5	Sequence 5, Appli
42	79.5	4.2	831	2	US-08-484-950-5	Sequence 5, Appli
43	79.5	4.2	831	2	US-08-757-653-5	Sequence 5, Appli
44	79.5	4.2	831	2	US-08-599-491-5	Sequence 5, Appli
45	79.5	4.2	831	2	US-08-756-386-5	Sequence 5, Appli

ALIGNMENTS

RESULT 1
US-08-613-220B-4
; Sequence 4, Application US/08613220B
; Patent No. 5958751
; GENERAL INFORMATION:
; APPLICANT: Murphy, Dennis
; APPLICANT: Reid, John
; TITLE OF INVENTION: ALPHA-GALACTOSIDASE
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson, P.C.
; STREET: 4225 Executive Square, Suite 1400
; CITY: La Jolla
; STATE: CA
; COUNTRY: US
; ZIP: 92037
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: Windows95
; SOFTWARE: FASTSEQ for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/613,220B
; FILING DATE: 08-MAR-1996
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Haile, Ph.D., Lisa A.
; REGISTRATION NUMBER: 38,347
; REFERENCE/DOCKET NUMBER: 09010/004001
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 619-678-5070
; TELEFAX: 619-68-5099
; TELEX:
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 346 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FRAGMENT TYPE: internal
US-08-613-220B-4

Query Match 84.2%; Score 1580; DB 2; Length 346;
Best Local Similarity 94.5%; Pred. No. 6.7e-159;
Matches 344; Conservative 1; Mismatches 1; Indels 18; Caps 18;

Qy 1 LRALVFNGLQYAEIPKSEIPKVIKAYIPVIETLIKKEIPGLNITGYTLKFLPKDIID 60
Db 1 LRALVFNGLQYAEIPKSE-PKVIKAYIPVIETLIKKEE-PGLNITGYTLKFLPKDIID- 57
Qy 61 LVKGGIASDIEIGTSYTHAILPLPLSRVQAQVORDREVKEELFELSPKGFWLPYLA 120
Db 58 LVKGGIASDIEIGTSYTHAILPLPLSRVQAQVORDREVKEELFELSPKGFWLPYLA- 114
Qy 121 DPTIIPALIKDNGYEYLFADGEAMLSAHLNSAIKPKIPLYPHLIKAQREKRYISYLLG 180
Db 115 DPTIIPALIKDNGYEYLFAD-EAMLSAHLNSAIKPKIPLYPHLIKAQREKRYISYLL- 171
Qy 181 LREURKAIVFEGKVTLKAVKIDIEAVPVVAVNTAVMLGIGRLPLMNPVKVSWIEBK 240
Db 172 LRELURKAIVFEGKVTLK-VKIDIEAVPVVAVNTAVML-IGRLPLMNPVKVSWIEBK- 228
Qy 241 NILLYGTDIEFIGYDIAGYRMSVEGLEVIDELNSELCLPSELKHSRGRELYLRTSSNA 300
Db 229 NILLYGTDIEFIGYDIAG-RMSVEGLEVIDELNSELCLPSELKHSRGRELYLRTSSNA- 285
Qy 301 DKSURIWREDEGNARLNLNMYNNRGEALALAEASDARGWEPLPERRLDAPRAIYNDRGE 360
Db 286 DKSURIWREDEGNARLNLNMYNNRGEALALAEASDARGWEPLPERRLDAPRAIYNDRGE- 342
Qy 361 NGEP 364
Db 343 NGEP 346
RESULT 2
US-07-894-212A-8
; Sequence 8, Application US/07894212A
; Patent No. 5366883
; GENERAL INFORMATION:
; APPLICANT: ASADA, KIYOZO
; APPLICANT: UEMORI, TAKASHI
; APPLICANT: MURAI, HIROYUKI
; APPLICANT: KATO, IKUNOSHIN
; APPLICANT: LADERMAN, KENNETH
; APPLICANT: ANFINSEN, CHRISTIAN
; TITLE OF INVENTION: THE ALPHA-AMYLASE GENE
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CUSHMAN DARB & CUSHMAN
; STREET: 1100 NEW YORK AVENUE, N.W.
; CITY: WASHINGTON, D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/894,212A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: KOKULIS, PAUL N.
; REGISTRATION NUMBER: 16773
; REFERENCE/DOCKET NUMBER: 95469/C-1195
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 861-3000
; TELEFAX: (202) 822-0944
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 647 amino acids
; TYPE: amino acid
; STRANDEDNESS: Single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-07-894-212A-8

Query Match 7.5%; Score 141.5; DB 1; Length 647;
Best Local Similarity 22.8%; Pred No. 3.4e-06;
Matches 89; Conservative 60; Mismatches 127; Indels 115; Gaps 23;
Qy 25 EKAYIPVIETLIKKEIP---FGLNITGYTLKELP---KDIIDLKGGIADLIEIGTSY 78
Db 27 EKCYWPFLETL--EYENPMKVAIHTSGPLIEWLQDNRRPEYIDLLRSYKRGQVEIVWAGE 84
Qy 79 THAILPLELSRVEAQVORDREVKE--ELFELSPKGFWLPYLAQVPIIPAILKONGEYL 136
Db 85 YEPVLASIP---KEDIEQIRLMKEWAKSIGDARGVWLTERVWQPELVKTLKESGIDYV 141
Qy 137 FADGEAMLSAHLNSAIKPKIPLY-PHLIKAQRE-----KRFYISYLLGLRELKKA 187
Db 142 IVD-----DYHMSAGLSKEELYWNPYTTEDGGEVIAVPPIDEKLR---YLIPFPVKV 192
Qy 188 IKL---VFEKGVTLKAV--KDIEAVPVVAVNTAVMLGIGRLPLMNPVKVASWI----- 236
Db 193 LEYLHSLIDGDESKVAVFHDGGEKFGIWFGEYEWY-----EKGLREFPDR 239
Qy 237 ---BDKNILLYGTDIE-----FIGYDIAGYRMSVEGLEVIDELNSELCLP----- 281
Db 240 ISSDEKINMLYTEILEKYKPRGLVYPIASY-----FEM-----SEWSLPKQARLF 287
Qy 282 ---SELKHSG---RELILTSSWAPDKSLRIWREDEGN---ARLNLNMYNNRGEALAL 330
Db 288 VEFVNEKLVKGIKFEYKRVFVGGI---KNF-FYKTPESYNNHMKMLVSKLVRRN----- 339
Qy 331 AENSARGWEPLPERRLDAPRAIYNDRG 359
Db 340 -----PEARKYLLRAQCNDAYWHG 358

RESULT 3
US-07-894-212A-2
; Sequence 2, Application US/07894212A
; Patent No. 5366883
; GENERAL INFORMATION:
; APPLICANT: ASADA, KIYOZO
; APPLICANT: UEMORI, TAKASHI
; APPLICANT: MURAI, HIROYUKI
; APPLICANT: KATO, IKUNOSHIN
; APPLICANT: LADERMAN, KENNETH
; APPLICANT: ANFINSEN, CHRISTIAN
; TITLE OF INVENTION: THE ALPHA-AMYLASE GENE
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: CUSHMAN DARB & CUSHMAN
; STREET: 1100 NEW YORK AVENUE, N.W.
; CITY: WASHINGTON, D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/894,212A
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: KOKULIS, PAUL N.
; REGISTRATION NUMBER: 16773
; REFERENCE/DOCKET NUMBER: 95469/C-1195
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 861-3000
; TELEFAX: (202) 822-0944
; TELEX: 6714627 CUSH
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:

LENGTH: 643 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-894-212A-2

Query Match 7.58; Score 141.5; DB 1; Length 649;
Best Local Similarity 22.88; Pred. No. 3.4e-06;
Matches 89; Conservative 60; Mismatches 127; Indels 115; Gaps 23;

QY 25 EKAYIPVETLIKKEIP---FGLNITGYTLKFLP---KDIIDLKVGKGIASDLIEIIGTYS 78
DB 29 EKCYWPFLETL--EYPMKVAIHTSGPLIEWLQDNRPVIDLLRSVLKRGQVEIVVAGF 86
QY 79 THAILPLPLSRVAQVORDREVKE--ELFELSPKGFMLPELAYDPIIPAILKDNQVEYL 136
DB 87 YEPVLASIP---KEDRIEQIRLMKEMAKSIGFDARGVWLTERVWQPELVKTLKESGIDIV 143
QY 137 FADGEAMLFSAHLSAIPKIPLY--PHLIKAQRE-----KRFYVSYLLGLRELKRA 187
DB 144 IVD-----DYHFSAGLSKBEELWYPTEDGGEVIAVFPIDEKLR---YLIFRPVQKV 194
QY 188 IKL---VPEGKVTLKAV--KDIKAVPVWVAVNTAVMLGIGLPLMNPKKVASMI----- 236
DB 195 LEYLSLIDGDESKVAVPHDDGEGFGWPGYEWY-----EKGWLRFFDR 241
QY 237 ---EKKDNLILYGTDIE-----FIGYRDIAGYRMSVEGLLEVIDELNSELCLP----- 281
DB 242 ISSDEKINLMLEYLEKYKPRGLVPLIASY-----PEM-----SEWSLPAKQARLF 289
QY 282 ----SELKHSQ-----RELYLTSSWAPDKSLRWREDEGN---ARLNMLSYNNRGEALLL 330
DB 290 VEFVNELVKGIKIFERYVFRGGIN---KNF-FYKYPESNTMHRMLVSKLVNRRN---- 341
QY 331 AENSARGWEPLPERRLDAFRAIYND--WRG 359
DB 342 -----PEARKYLLRAQCNDAYWHG 360

RESULT 4

US-07-893-928A-1

Sequence 1, Application US/07893928A
Patent No. 5578479

GENERAL INFORMATION:

APPLICANT: LADERMAN, KENNETH
APPLICANT: ANFINSEN, CHRISTIAN
TITLE OF INVENTION: a-AMYLASE FROM HYPERTHERMOPHILIC
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: CUSHMAN DARBY & CUSHMAN
STREET: 1100 NEW YORK AVENUE, N.W.
CITY: WASHINGTON, D.C.
COUNTRY: U.S.A.
ZIP: 20005

COMPUTER READABLE FORM:

MEDIUM TYPE: Tape
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
FILING DATE:

APPLICATION NUMBER: US/07/893.928A

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:
NAME: KOKULIS, PAUL N.
REGISTRATION NUMBER: 16773
REFERENCE/DOCKET NUMBER: 95470/C-1197
TELECOMMUNICATION INFORMATION:
TELEPHONE: (202) 861-3000
TELEFAX: (202) 822-0944

TELEX: 6714627 CUSH
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 650 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-07-893-928A-1

Query Match 7.58; Score 141.5; DB 1; Length 650;
Best Local Similarity 22.88; Pred. No. 3.4e-06;
Matches 89; Conservative 60; Mismatches 127; Indels 115; Gaps 23;

QY 25 EKAYIPVETLIKKEIP---FGLNITGYTLKFLP---KDIIDLKVGKGIASDLIEIIGTYS 78
DB 29 EKCYWPFLETL--EYPMKVAIHTSGPLIEWLQDNRPVIDLLRSVLKRGQVEIVVAGF 86
QY 79 THAILPLPLSRVAQVORDREVKE--ELFELSPKGFMLPELAYDPIIPAILKDNQVEYL 136
DB 87 YEPVLASIP---KEDRIEQIRLMKEMAKSIGFDARGVWLTERVWQPELVKTLKESGIDIV 143
QY 137 FADGEAMLFSAHLSAIPKIPLY--PHLIKAQRE-----KRFYVSYLLGLRELKRA 187
DB 144 IVD-----DYHFSAGLSKBEELWYPTEDGGEVIAVFPIDEKLR---YLIFRPVQKV 194
QY 188 IKL---VPEGKVTLKAV--KDIKAVPVWVAVNTAVMLGIGLPLMNPKKVASMI----- 236
DB 195 LEYLSLIDGDESKVAVPHDDGEGFGWPGYEWY-----EKGWLRFFDR 241
QY 237 ---EKKDNLILYGTDIE-----FIGYRDIAGYRMSVEGLLEVIDELNSELCLP----- 281
DB 242 ISSDEKINLMLEYLEKYKPRGLVPLIASY-----PEM-----SEWSLPAKQARLF 289
QY 282 ----SELKHSQ-----RELYLTSSWAPDKSLRWREDEGN---ARLNMLSYNNRGEALLL 330
DB 290 VEFVNELVKGIKIFERYVFRGGIN---KNF-FYKYPESNTMHRMLVSKLVNRRN---- 341
QY 331 AENSARGWEPLPERRLDAFRAIYND--WRG 359
DB 342 -----PEARKYLLRAQCNDAYWHG 360

RESULT 5

US-08-270-013B-2

Sequence 2, Application US/08270013B
Patent No. 5686294

GENERAL INFORMATION:

APPLICANT: Sogabe et al.
TITLE OF INVENTION: PROTEIN HAVING HEAT-RESISTANT MALATE
TITLE OF INVENTION: DEHYDROGENASE ACTIVITY
NUMBER OF SEQUENCES: 2
CORRESPONDENCE ADDRESS:
ADDRESSEE: Leydig, Voit & Mayer, Ltd.
STREET: Two Prudential Plaza, Suite 4900
CITY: Chicago
STATE: Illinois
COUNTRY: USA
ZIP: 61601-6780

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/270.013B
FILING DATE: 01-JUL-1994
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: JP 164701/1993
FILING DATE: 02-JUL-1993
ATTORNEY/AGENT INFORMATION:

```
; NAME: Green, Robert F.
; REGISTRATION NUMBER: 27555
; REFERENCE/DOCKET NUMBER: 62321
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 616-5600
; TELEFAX: (312) 616-5700
; TELEX: (25)3533
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 329 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-270-013B-2

Query Match 5.58; Score 104; DB 1; Length 329;
Best Local Similarity 26.08; Pred. No. 0.011;
Matches 40; Conservative 24; Mismatches 50; Indels 40; Gaps 7;

QY 10 LQYAEIPKSEIPK-----VIEKAYIPVETLKEEIPFGLNITGYTLKFLPKDIIDL 61
DB 130 MTYTVFKESGFPKRVIGSQSLDTR---FRTFVAEE-----LNIS-----VKDVTGF 175
QY 62 VKGGTASDLIEIGTSYTHAI--LPPLPSRVEAQVORDREVKEELFELSPKG--FWLPE 117
DB 176 VLGGHGDMDVPLVRSYAGGIPLEKLPKDRDLDAIVERTKGGGEIVNLLGNSAYYAPA 235
QY 118 LAYDPIIPAILKDN-----GVEYLF 137
DB 236 ASLVEMVEAILKQDRRLPAIAYLEGEVGYEGYI 269

RESULT 6
US-08-838-418-2
; Sequence 2, Application US/08838418
; Patent No. 5744342
; GENERAL INFORMATION:
; APPLICANT: Sogabe et al.
; TITLE OF INVENTION: PROTEIN HAVING HEAT-RESISTANT MALATE
; TITLE OF INVENTION: DEHYDROGENASE ACTIVITY
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Leydig, Voit & Mayer, Ltd.
; STREET: Two Prudential Plaza, Suite 4900
; CITY: Chicago
; STATE: Illinois
; COUNTRY: USA
; ZIP: 60601-6780
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/838,418
; FILING DATE: 17-MAR-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 164701/1993
; FILING DATE: 02-JUL-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Hoover, Allen E.
; REGISTRATION NUMBER: 37354
; REFERENCE/DOCKET NUMBER: 78339
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (312) 616-5600
; TELEFAX: (312) 616-5700
; TELEX: (25)3533
```

```
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 329 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-838-418-2

Query Match 5.58; Score 104; DB 1; Length 329;
Best Local Similarity 26.08; Pred. No. 0.011;
Matches 40; Conservative 24; Mismatches 50; Indels 40; Gaps 7;

QY 10 LQYAEIPKSEIPK-----VIEKAYIPVETLKEEIPFGLNITGYTLKFLPKDIIDL 61
DB 130 MTYTVFKESGFPKRVIGSQSLDTR---FRTFVAEE-----LNIS-----VKDVTGF 175
QY 62 VKGGTASDLIEIGTSYTHAI--LPPLPSRVEAQVORDREVKEELFELSPKG--FWLPE 117
DB 176 VLGGHGDMDVPLVRSYAGGIPLEKLPKDRDLDAIVERTKGGGEIVNLLGNSAYYAPA 235
QY 118 LAYDPIIPAILKDN-----GVEYLF 137
DB 236 ASLVEMVEAILKQDRRLPAIAYLEGEVGYEGYI 269

RESULT 7
US-08-418-893D-23
; Sequence 23, Application US/08418893D
; Patent No. 5559220
; GENERAL INFORMATION:
; APPLICANT: ROESSLER, PAUL G
; APPLICANT: OHLROGGE, JOHN B
; TITLE OF INVENTION: GENE THAT ENCODES ACETYL-COENZYME A
; TITLE OF INVENTION: CARBOXYLASE FROM CYCLOTHELLA CRYPTICA
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NATIONAL RENEWABLE ENERGY LABORATORY
; STREET: 1617 Cole Blvd.
; CITY: Golden
; STATE: CO
; COUNTRY: USA
; ZIP: 80401-3393
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent in Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/418,893D
; FILING DATE: April 7, 1995
; CLASSIFICATION: 800
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/104,938
; FILING DATE: September 14, 1993
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: O'CONNOR, EDNA
; REGISTRATION NUMBER: 29,252
; REFERENCE/DOCKET NUMBER: MRI/NREL IR# 92-48CON
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 303-231-1000
; TELEFAX: 303-231-1098
; TELEX:
; INFORMATION FOR SEQ ID NO: 23:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2089 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
```

US-08-418-893D-23

Query Match 4.7%; Score 88.5; DB 1; Length 2089;
Best Local Similarity 19.6%; Pred. No. 9.2;
Matches 70; Conservative 65; Mismatches 110; Indels 113; Gaps 21;

QY 9 NLQYAEIPKSEIPKVEKAVIPVETLIK--EEIP-----FGLNITGVTLPKPLKDIIDL 61
DB 1057 SFQFADVPKAD--RVTRQGFSSVIDDASKFAQOLPELNSFGSKIAGDASKEGPNVQLQV 1114
QY 62 VKGGIASDL-IEII--GTSYTHAILPLLSRVEAQVQDRVKEELFELSPKGFWLPEL 118
DB 1115 --GALSGDISIEDLEKATSANKDKLNLGVRTVTALIPRKK-----DPSYYSFFQC 1164
QY 119 A---YDPIIPAILKDCNGEYFLPADGEAMLFSAHLSAIPKIPLYPHLIKAOR-EKRF-- 172
DB 1165 SGFKEDPL-----RGMRTFHHLELGRLEENFAL 1195
QY 173 -----RYISYLLGLREL--RKAIKLVFEGKVTLKAVKDIEAVPVWAVNTAVMLGIGR 223
DB 1196 ERIPAVGRNVQIYVSGSEKTARRNAQVVF-----LRAISHTPGLTTFSGARRALLQGLDE 1250
QY 224 LPLNPKKVASWIEDKDNILLYG-----TDI-----EFIGYRDIAGYRMSVEGLEVID 272
DB 1251 LERAQANSKVS-VQSSSRIYLSLPEQSDATPEEIAKEFE-----VID 1293
QY 273 ELNSELCLPSELKHSRGELYRTSSWAPDKSLRIWREDE-GNAR---LNLMSYNNMRGE 326
DB 1294 KLKSR-----AQLTKLRVDE--IETKRVTVQDEDEGSPRVVPRVLRVASSMQGE 1341

RESULT 8

US-08-418-893D-24
; Sequence 24, Application US/08418893D
; Patent No. 5559220

; GENERAL INFORMATION:
; APPLICANT: ROESSLER, PAUL G
; APPLICANT: OHLEGG, JOHN B
; TITLE OF INVENTION: GENE THAT ENCODES ACETYL-COENZYME A
; TITLE OF INVENTION: CARBOXYLASE FROM CYCLOTELLA CRYPTICA
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: NATIONAL RENEWABLE ENERGY LABORATORY
; STREET: 1617 Cole Blvd.
; CITY: Golden
; STATE: CO
; COUNTRY: USA
; ZIP: 80401-3393

; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent In Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/418,893D
; FILING DATE: April 7, 1995
; CLASSIFICATION: 800

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/104,938
; FILING DATE: September 14, 1993
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: O'CONNOR, EDNA
; REGISTRATION NUMBER: 29,252
; REFERENCE/DOCKET NUMBER: MRI/NREL IR# 92-48CON
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 303-231-1000
; TELEFAX: 303-231-1098
; TELEX:

; INFORMATION FOR SEQ ID NO: 24:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2089 amino acids

; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
; FRAGMENT TYPE: N-terminal
US-08-418-893D-24

Query Match 4.7%; Score 88.5; DB 1; Length 2089;
Best Local Similarity 19.6%; Pred. No. 9.2;
Matches 70; Conservative 65; Mismatches 110; Indels 113; Gaps 21;

QY 9 NLQYAEIPKSEIPKVEKAVIPVETLIK--EEIP-----FGLNITGVTLPKPLKDIIDL 61
DB 1057 SFQFADVPKAD--RVTRQGFSSVIDDASKFAQOLPELNSFGSKIAGDASKEGPNVQLQV 1114
QY 62 VKGGIASDL-IEII--GTSYTHAILPLLSRVEAQVQDRVKEELFELSPKGFWLPEL 118
DB 1115 --GALSGDISIEDLEKATSANKDKLNLGVRTVTALIPRKK-----DPSYYSFFQC 1164
QY 119 A---YDPIIPAILKDCNGEYFLPADGEAMLFSAHLSAIPKIPLYPHLIKAOR-EKRF-- 172
DB 1165 SGFKEDPL-----RGMRTFHHLELGRLEENFAL 1195
QY 173 -----RYISYLLGLREL--RKAIKLVFEGKVTLKAVKDIEAVPVWAVNTAVMLGIGR 223
DB 1196 ERIPAVGRNVQIYVSGSEKTARRNAQVVF-----LRAISHTPGLTTFSGARRALLQGLDE 1250
QY 224 LPLNPKKVASWIEDKDNILLYG-----TDI-----EFIGYRDIAGYRMSVEGLEVID 272
DB 1251 LERAQANSKVS-VQSSSRIYLSLPEQSDATPEEIAKEFE-----VID 1293
QY 273 ELNSELCLPSELKHSRGELYRTSSWAPDKSLRIWREDE-GNAR---LNLMSYNNMRGE 326
DB 1294 KLKSR-----AQLTKLRVDE--IETKRVTVQDEDEGSPRVVPRVLRVASSMQGE 1341

RESULT 9

US-09-141-135-2
; Sequence 2, Application US/09141135
; Patent No. 5981729

; GENERAL INFORMATION:
; APPLICANT: CHUN, Jong Yoon
; APPLICANT: LEE, Jong Hun

; TITLE OF INVENTION: Transcription Factor Gene Induced by Water Deficit and Abscisi
; TITLE OF INVENTION: Acid Isolated from Arabidopsis thaliana

; FILE REFERENCE: 1942/31

; CURRENT APPLICATION NUMBER: US/09/141,135

; CURRENT FILING DATE: 1998-08-27

; NUMBER OF SEQ ID NOS: 2

; SOFTWARE: WordPerfect 6.1/Windows

; SEQ ID NO 2

; LENGTH: 235

; TYPE: PRT

; ORGANISM: Arabidopsis thaliana

US-09-141-135-2

Query Match 4.6%; Score 87; DB 2; Length 235;

Best Local Similarity 21.0%; Pred. No. 0.39;

Matches 49; Conservative 37; Mismatches 71; Indels 76; Gaps 11;

QY 165 KAOREKFRYISYLLGLRELKKAIKLVFEGKVTLKAVKDIEAVPVWAVNTAVMLGIGRL 224
DB 27 KSNOKRFN-----EEQKSLLELIFESETRLEPRKKVQ-----VARELG----- 65
QY 225 PLMPKPKVASWIEDK-----NILLYGTDTIEFIGYRDIAG-----YRMSVEGLL 268
DB 66 --LQPRMTIWFQNKRAFWTKOLEKEYNLIR-----ANYNNLASQFEIMKKEQSLV 116
QY 269 EVIDELNSELCLPSELKH-----SGRELYLRTSSWAPD-KSLRIWREDEGNARLNLMSYN- 322

```
Db 117 SELORNEEMOREKKEHCECGDGLALSSSTESHNGKSEPEGRDQGSVLCNDGDYNN 176
Qy 323 -----MEG-----ELALLAENSARGNEPLPERRLDAFRAIYNDRGENGE 363
Db 177 NIKTEYFRVOGETDHELMNIVEKAD-----DSCLTSSENWGGFNSD 217

RESULT 10
US-09-355-166-1
; Sequence 1, Application US/09355166
; Patent No. 6316241
; GENERAL INFORMATION:
; APPLICANT: Genencor International, Inc.
; TITLE OF INVENTION: Alpha/Beta Hydrolase-Fold Enzymes
; FILE REFERENCE: GC511-PCT
; CURRENT APPLICATION NUMBER: US/09/355,166
; CURRENT FILING DATE: 1999-07-20
; NUMBER OF SEQ ID NOS: 21
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 657
; TYPE: PRT
; ORGANISM: Bacillus
US-09-355-166-1
```

```
Query Match 4.6%; Score 87; DB 4; Length 657;
Best Local Similarity 22.4%; Pred. No. 2.1; Mismatches 106; Indels 88; Gaps 17;
Matches 70; Conservative 49;

Qy 48 GYTLKFLPK---DIIDL-----VKGSIASOLIEI-IGTSYT-----HAILPLPLSRVEA 93
Db 75 GRTLAFISDREGDAAQLYIMSTEGEARKLTDIPYGVSKPLWSPOGESILVTSLGEGES 134
Qy 94 QVQDREVKE-----ELFELSPK--GFWLPELAYDPIIPAILK-----DNGYE 134
Db 135 IDREKTEQDSYEPVEQVUSYRDKGGLTRGAYAOVLVLVSVKSGEMKELTSHKHADHGP 194
Qy 135 YLFADGAMLFSAHL-----NSAIRP-----IKLYPH-----LIKAQREKFRVIS 176
Db 195 AFSPDGKWLVSANLTETDASKPHDVYIMSLSEGLKQVTPHRGSGSSSPDGRVLA 254
Qy 177 YLLGLRLRA-----IKLVFGKVT--LKAVKDIEAVPVWAVNTAVMLGRLPLMNP 229
Db 255 LLGNEKEYKNALSKAWLYDIEQGRJLTLEMLDVHAD--ALIGSLGGAEQRP----- 309
Qy 230 KKVASWIEDKDNILLYGTDTIEFGYRDIAGYRMSVGLLEVIDELNSELCPLSELKHSGR 289
Db 310 -----WTKDSQGFVVIGTDQGSTGI-----YYSIBGLVYPI-RLEKE----- 346
Qy 290 ELYLRTSSWAPDK 302
Db 347 --YINFSLSLSPDE 357
```

```
RESULT 11
US-08-248-021A-2
; Sequence 2, Application US/08248021A
; Patent No. 5648240
; GENERAL INFORMATION:
; APPLICANT: Hook, Magnus
; APPLICANT: Jonsson, Klas
; APPLICANT: Patti, Joseph M.
; APPLICANT: Gurusidappa, Sivashankarappa
; TITLE OF INVENTION: MHC II ANALOG FROM STAPHYLOCOCCUS AUREUS
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P.O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: United States of America
```

```
ZIP: 77210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/248,021A
FILING DATE: 24-MAY-1995
CLASSIFICATION: 424
ATTORNEY/AGENT INFORMATION:
NAME: Kitchell, Barbara S.
REGISTRATION NUMBER: 33,928
REFERENCE/DOCKET NUMBER: TANK:155
TELECOMMUNICATION INFORMATION:
TELEPHONE: (512) 418-3000
TELEFAX: (512) 474-7577
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 689 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: linear
US-08-248-021A-2
```

```
Query Match 4.6%; Score 86.5; DB 1; Length 689;
Best Local Similarity 18.8%; Pred. No. 2.5;
Matches 51; Conservative 38; Mismatches 79; Indels 103; Gaps 11;

Qy 33 ETLIKSEIPEGLNITGYTLKFLP-----KDIIDLKVGSIASD----- 69
Db 264 EAKVNNQVPSYINLNGTTTNIOSNLAFAFNKPTNKNLTTRKVSCLKSDRGVSEDLKHA 323
Qy 70 -----LIEIIGTSYTHAILPLPLSRVEAQVQDREVKEEL---FELSPK 111
Db 324 KKAYTYVYFKNGCKRVHLSNNTYNTANLVHAKDVKRIEVTVTSKVKAERYVPTIAVN 383
Qy 112 GFWLPEI-----AYDPT---IPAILK-DNG-----YELFADGAM 143
Db 384 GASNPITLSDLKFTGDSRVSYSDIKKKVSVLKHDRGIGERELKYAEKATYTVHFKNGTK 443
Qy 144 LFSAHLSAIPKIPLYPHLIK-----AOREKFRVIS----- 177
Db 444 VI--NLNSNISQNLNLYVRDKNIDIDVKTGAKAKVYSVYPYTIANGTTTPIASKLXL 501
Qy 178 ---LLGLRELKRAIKLVFEGKVTLKAVKDIE 205
Db 502 NKQLIGYQDLNKKVSKVIKHD---RGINDIE 529
```

```
RESULT 12
US-09-235-451-36
; Sequence 36, Application US/09235451
; GENERAL INFORMATION:
; APPLICANT: Julius, David J.
; APPLICANT: Caterina, Michael J.
; APPLICANT: Brake, Anthony J.
; TITLE OF INVENTION: NUCLEIC ACID SEQUENCES ENCODING
; TITLE OF INVENTION: CAPSAICIN RECEPTOR AND CAPSAICIN RECEPTOR-RELATED
; TITLE OF INVENTION: POLYPEPTIDES AND USES THEREOF
; FILE REFERENCE: 9076/084CIP
; CURRENT APPLICATION NUMBER: US/09/235,451
; CURRENT FILING DATE: 1999-01-22
; PRIOR APPLICATION NUMBER: 60/072,151
; PRIOR FILING DATE: 1998-01-22
; PRIOR APPLICATION NUMBER: 08/915,461
; PRIOR FILING DATE: 1997-08-20
; NUMBER OF SEQ ID NOS: 48
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 36
; LENGTH: 764
; TYPE: PRT
```

```

Query Match      4.48; Score 83.5; DB 1; Length 641;
Best Local Similarity 19.6%; Pred. No. 4.6;
Matches 73; Conservative 53; Mismatches 143; Indels 103; Gaps 16;

Qy 24 IEAXYIPVETIIKEBIPPEGLNAYITLKFPIKDIIDLVRKGIASDLI-----EIIGTSTV 79
      :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 116 LQXLQPTIIFINK-----IDRAGVNLRLYLDOI-----KANLSQDYLQMNVYDGSVP 165

Qy 80 HAILPLPLSRVEAQVQRDREYKEELF--ELSPKGFWLPCLAYDPIIPAILKXNDGYEVL 136
      :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 166 VCSQTYIKBEEYFCVNDHNDILERYLADSLSPADYW-----NTIILVAKAYVPVL 219

Qy 137 FADGEAMFLPSAHLNSAIKPI-----KPYPHIILKAQREKRPRTYSYLLGLBELR 185
      :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 220 --HGSAM-FNIGINEULLDAITFILPPASVSNRLSSVLYKIEHPKPGHKRSFL----- 269

Qy 186 KAIKLVEGKVTILKAVKDIEAVPVVAVNTAVMLGIGRLPLMNPKKVASIEDKDNILLY 245
      :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 270 ---KIIDGSLRLDRVVRIINDSEKFIKIKNLKTIINOGR----- 303

Qy 246 GTDIEFTGYRDIAGYRMSVGEGLLEWIDE-----LNSELCLPSELKHSGRELYLRTSSWA 299
      :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 304 EINDEVGANDIA-----IVEDMDDFRIGNYLGAEPCLIQGLSHQHAPAL---KSSVR 352

Qy 300 PDKSLR-----IWREDEGCRARLNMLSYNMRGELALLAENSARGVEPUPER----- 345

```

```
Db 353 PDRPEERSKVISALNLTWIEDP-SLSFSINSYSDELEISLYGLTKQKEIIQTLLERFSVK 411
QY 346 -RLDAFRAIYND 356
Db 412 VHFDEIKTIYKE 423

RESULT 15
US-08-161-999-3
: Sequence 3, Application US/08161999
: Patent No. 5674733
: GENERAL INFORMATION:
: APPLICANT: Salyers, Abigail A.,
: APPLICANT: Shoemaker, Nadja B.,
: APPLICANT: Nikolich, Mikeljon P.
: TITLE OF INVENTION: Method and Materials For
: TITLE OF INVENTION: Introducing DNA Into Prevotella ruminicola
: NUMBER OF SEQUENCES: 5
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: William Brinks Olds Hofer Gilson and Lione
: STREET: P.O. Box 10395
: CITY: Chicago
: STATE: Illinois
: COUNTRY: USA
: ZIP: 60610
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette, 5.25 inch, 360 kb storage
: COMPUTER: IBM XT
: OPERATING SYSTEM: MS-DOS 3.31
: SOFTWARE: WordPerfect 5.1
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/161,999
: FILING DATE: 02-DEC-1993
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 07/718,535
: FILING DATE: 05-JUN-1991
: ATTORNEY/AGENT INFORMATION:
: NAME: Crook, Wannell M.
: REGISTRATION NUMBER: 31,071
: REFERENCE/DOCKET NUMBER: 3617/22
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (312) 321-4200
: TELEFAX: (312) 321-4299
: INFORMATION FOR SEQ ID NO: 3:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 641 amino acids
: TYPE: amino acid
: STRANDEDNESS:
: TOPOLOGY: unknown
: US-08-161-999-3
```

```
Query Match 4.4%; Score 83.5; DB 1; Length 641;
Best Local Similarity 19.6%; Pred. No. 4.6;
Matches 73; Conservative 53; Mismatches 143; Indels 103; Gaps 16;

QY 24 IEKAYIPVIEFLKEIPFGLNITGYTLKPLKDIIDLKVGGIASDLI-----EIIGTSTY 79
Db 116 LQKLIQPTIIFINK-----IDRAGVNLERLYLDI-----KANLSQDVLEMQNVVDGSVYP 165
QY 80 HAILPLPLSRVEAQVORDREVKEELF---ELSPKGFWLPPELAYDPIIAILKDNCGEYL 136
Db 166 VCSQTYKEEYKEVCHNDNILERYLADSEISPADYV-----NTIIALVAKAKYIPVL 219
QY 137 FADGEAMLFSAHLSAIIKPI-----KPLYPHLIKAQREKFRYISYLLGLREL 185
Db 220 --HGSAM-FNIGINELLDATISFTLPPASVSNRLSSVLYKIEHDPKHKRSEL-----269
QY 186 KAIKLVFEGKVTILKAVKDIEAPVAVNTAVMLGTRPLMNPKKVASKWIEDKNILLY 245
Db 270 ----KIIDGSLRLRDVVRINDSEKFIKIKLKTINQGR-----303
```

```
QY 246 GTDIEFIFGYRDIAGYRMSVEGLLEVIDE-----INSELCLPSELKHSRGRELYLRTSSWA 299
Db 304 EINVDEVGANDIA-----IVEDMDDFRIGNYLGAEPCLIQGLSHQHPAL---KSSVR 352
QY 300 PKSLR-----IWREDESGNARLNMLSYMMRGELALLAENSARGWEPLPER-----345
Db 353 PDRPEERSKVISALNLTWIEDP-SLSFSINSYSDELEISLYGLTKQKEIIQTLLERFSVK 411
QY 346 -RLDAFRAIYND 356
Db 412 VHFDEIKTIYKE 423
```

Search completed: June 2, 2002, 18:06:15
Job time: 233 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: June 2, 2002, 18:03:43 ; Search time 20.94 Seconds
(without alignments)
1670.321 Million cells

US-09-886-400-4

• 0250,1203
: 2709 1201124

Sequence: 1 LRALVFHGNLQYAEIPKSEI.....RRLLDAFRAIYNDWRGENGEP 364

scoring table: BLOSUM62

SCOTTING CADRE: BLOSSOM2
Gapex 10.0 : Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0

Minimum DB seq length:	0
Maximum DB seq length:	2000000000

Post-processing: Minimum Match 0%

Post-processing:	Minimum	Match	100%
	Maximum	Match	100%

Maximum Match	100%
1st place	150,000,000
2nd place	100,000,000
3rd place	75,000,000
4th place	50,000,000
5th place	25,000,000
6th place	10,000,000
7th place	5,000,000
8th place	2,500,000
9th place	1,250,000
10th place	625,000
11th place	312,500
12th place	156,250
13th place	78,125
14th place	39,062
15th place	19,531
16th place	9,766
17th place	4,883
18th place	2,441
19th place	1,220
20th place	610
21st place	305
22nd place	152
23rd place	76
24th place	38
25th place	19
26th place	9
27th place	4
28th place	2
29th place	1
30th place	0

Детские книги

Database : PIR_71:*

```
l: pid:*
```

2: pir2:*

```
3: pir3:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query		DB	ID	Description
		Match	Length			
1	1494	79.6	364	2	E71144	hypothetical prote
2	169	9.0	529	2	AC2112	hypothetical prote
3	162.5	8.7	529	2	S76831	hypothetical prote
4	161.5	8.6	633	2	G71241	probable alpha-amylase A (EC 3.2
5	157.5	8.4	686	1	ADYAT	uncharacterized co
6	147	7.8	527	2	F97197	alpha-amylase (or
7	145.5	7.8	655	2	E75206	alpha-amylase (EC
8	141.5	7.5	467	2	B64501	alpha-amylase (EC
9	141.5	7.5	649	2	A49512	hypothetical prote
10	130.5	7.0	560	2	D71011	hypothetical prote
11	128	6.8	923	2	G83826	amylomullulanase P
12	124	6.6	1362	2	AC75207	hypothetical prote
13	123.5	6.6	1744	2	AC1970	methanol dehydroge
14	120.5	6.4	324	1	B69553	hypothetical prote
15	119	6.3	602	2	F75120	hypothetical prote
16	113	6.0	1069	2	AF3930	malate dehydrogena
17	112	6.0	312	2	D40383	conserved hypothet
18	111	5.9	526	2	D71334	alpha amylase [imp
19	109.5	5.8	447	2	E90250	hypothetical prote
20	106	5.6	526	2	B70859	conserved hypothet
21	105.5	5.6	902	2	E90270	hypothetical prote
22	104	5.5	704	2	G95587	malate dehydrogena
23	103.5	5.5	315	2	F84044	vacu protein (ribo
24	101.5	5.4	705	2	F70475	probable xylanase/
25	100.5	5.4	266	2	G96943	probable alpha-amylase
26	100.5	5.4	619	2	D71361	BIR repeat contain
27	99	5.3	4845	2	T31067	two-component resp
28	98.5	5.2	588	2	AE1926	hypothetical prote
29	98	5.2	503	2	T19319	

ALIGNMENTS

```

RESULT      1
E71144     hypothetical protein PH0368 - Pyrococcus horikoshii
C:Species: Pyrococcus horikoshii
C:Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 08-Sep-2000
C:Accession: E71144
R:Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Haikawa, Y.; Hino, Y.; Yamamoto, S.; Se
M.: Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Ogu
DNA Res. 5, 55-76, 1998
A:Title: Complete sequence and gene organization of the genome of a hyper-thermophilic
A:Reference number: A71000; WUID:98344137
A:Accession: E71144
A>Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1364 <RAW>
A:Cross-references: GB:AP000002; NID:g3236129; PIDN:BAA29442.1; PID:d1030385; PID:g32
A:Experimental source: strain OT3
A>Note: This accession replaces an interim accession for a sequence replaced by GenBa
C:Genetics:
A:Gene: PH0368
C:Superfamily: Pyrococcus horikoshii hypothetical protein PH0368

Query Match          79.6%; Score 1494; DB 2; Length 364;
Best Local Similarity 75.8%; Pred. No. 2.7e-105;
Matches 275; Conservative 49; Mismatches 39; Indels 0; Gaps 0;

Qy   1 LRALVFHGNLQYAEIPKSETPKVIEAKIPIVETLIKKEIPFGLNITGYLKFLPKDIID 60
Db   1 MRALIFHGNIQYAEIPKHEISKVIEKSIFYFTISELIKREIPFGNITGYSIQLFQPELIH 60

Qy   61 LVKGTIASDLTEIGTSTHTAILPLLSRVQAQVRDREVKEELFELSFKGWLPPLAY 120
Db   61 LIKREGISELTIELGTSTHTAILPILLSRTREAQIKRDRIKEIFEVSPPGGFWLPPLAY 120

Qy   121 DPTIPAILKONGYEYLFDAGFAMLLFSAHLSNAIKPIKPILYPHLTKAQEKRRFYISVLG 180
Db   121 DPTIIPAILRDNEYLYLFDAGGEAMLLFNHLNSAIKSIAPLYLKAQQRGGFVYNLVLG 180

Qy   181 LRLEUKAIKLVEFGKVTALKAVKDIEAYPVWVVNTATVMGIQRLPLMKPKVASWIEDKD 240
Db   181 LRCLKAINLTFGSKVILEAVKDIEAIPVWVSINIAIMLGARFPPLMSPRXVANWIWKGD 240

Qy   241 NILLYGTDTDFIGVRTAGVMSVEGLEVIDELNSLCPLSELKHSGRELILTSSSWAP 300
Db   241 ELILYGTDIEFLGYRSIAGHKIASGLMEVLNLGEGELCPDRIRHNGRRLDYLRATSSWAP 300

Qy   301 DKSLRWREDGNARNLMNVSYMGELLALAENSARDARGWEPLPERRLDAFRATVNDWRGE 360
Db   301 DKSLRWKEDEGNARNLMNYICMDGEEAFLAENSDARGWEPLPERLDATKATKYWNNE 360

Qy   361 NCE 363

```

C:Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 20-Jun-2000
C:Accession: S76831
R:Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, O. K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yas
DNA Res. 3, 109-136, 1996
A:Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechocys*
S.
A:Reference number: S74322; MUID:97061201
A:Accession: S76831
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-529 <KAN>
A:Cross-references: EMBL:D0916; GB:AB001339; NID:gi653715; PIDN:BAA18743.1; PID:gi65
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
C:Superfamily: *Pyrococcus horikoshii* hypothetical protein PH1386

Query Match 8.7%; Score 162.5; DB 2; Length 529;
Best Local Similarity 21.4%; Pred. NO. 0.00017;
Matches 101; Conservative 47; Mismatches 140; Indels 183; Gaps 19;

QY 3 ALVPHGNLQVAEIPKSE-----IPKVIKAYIPV---ETLKEEIPFGLNIT----- 47
DB 7 ALVLHAHLPFVRHPESDYVLEEWLYEATITETIPIHVFEGKRGDGVDFKMTSMTPPL 66
QY 48 -----GYTLKFLPK----- 56

DB 67 VSMRLDPLQRYEADLALELLEGERNAQNGHLRYLAERHATFNEARQMWERYNG 126

QY 57 DIIDLKGGIASDLIEIGTSYTHAIPPLPL--SRVEAQVQDRVKKELFELSPKGF 114
DB 127 DLVTAFAKQFQDSNLEITTCGATHGYLPKMKYPOAWAQIYVACEHYEETTCRPPKGIW 186

QY 115 LPELAYDPIIPAILKONGEYELFADGAMLFSA-----HUNS 151
DB 187 LPECAVYEGLEMLADAGLRYFTDGHGILYARPRFGTYAPIFTGTGVAAFGRDHES 245

QY 152 -----AIPKIPPLPHLIK-AQREKRPYISYLLGLRELKAKLVF-----EGKVT 197

DB 247 QQVSSQVGPVGPVREFYKDLGWEAEYIYKPYIMPNGQRNKIGIKYHKTSDGGLS 306

QY 198 LKAVKDIEAVPVVAVVAVTAV-----MLGH-GRPLMNPKKVAS-----WI 236

DB 307 EKAWYD---PYNAKAKAEHASNEMVNRQVQVGLSGIMGRPPLVSPYDAELFGHWY 362

QY 237 E-----DKDNI-----LYGTDFEPI-----GYRDIAGY----- 260

DB 363 EGPWFIDYLFKRSWFDQDTFEMTHLADYLRGNPHQVCRPSOSSWGYKGFHEYWLNDTNA 422

QY 261 -----RMSVEGLLEVIDELSELCLPSELKHSGRELYL-RTSSWA 299

DB 423 WIYPHLHKAERMIELSHREAVDELEK-----ALNQAARELLLAQSSDWA 468

RESULT 4
G71241
probable alpha-amylase - *Pyrococcus horikoshii*
C:Species: *Pyrococcus horikoshii*
C:Date: 14-Aug-1998 #sequence_revision 14-Aug-1998 #text_change 21-Jul-2000
A:Accession: G71241
R:Kawarabayashi, Y.; Sawada, M.; Horikawa, H.; Hata, Y.; Yamamoto, S.; Se
M.; Ohfuku, Y.; Funahashi, T.; Tanaka, T.; Kudoh, Y.; Yamazaki, J.; Kushida, N.; Ogu
DNA Res. 5, 55-76, 1998
A:Title: Complete sequence and gene organization of the genome of a hyper-thermophil
A:Reference number: A71000; MUID:98344137
A:Accession: G71241
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-633 <KAW>
A:Cross-references: GB:AF000001; NID:g2326128; PIDN:BAA29262.1; PID:g3256579
A:Experimental source: strain OT3
A:Note: this accession replaces an interim accession for a sequence replaced by GenBa
C:Genetics:

Db 361 NGK 363

RESULT 2
AC2112
hypothetical protein alr2450 [imported] - *Anabaena* sp. (strain PCC 7120)
C:Species: *Anabaena* sp.
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 11-Jan-2002
C:Accession: AC2112
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriuchih
Nakazaki, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium *Ana*
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AC2112
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-529 <KOR>
A:Cross-references: GB:BA000019; PIDN:BA74149.1; PID:gl7131542; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: alr2450
C:Superfamily: *Pyrococcus horikoshii* hypothetical protein PH1386

Query Match 9.0%; Score 169; DB 2; Length 529;
Best Local Similarity 21.0%; Pred. NO. 5.6e-05;
Matches 104; Conservative 55; Mismatches 162; Indels 174; Gaps 21;

QY 3 ALVPHGNLQVAEIPKSE-----IPKVIKAYIPV---VIETLKEEIPFGL----- 44
DB 7 ALVLHAHLPFVRHPESDYVLEEWLYEATITETIPIKLVDFGLKRGIDFKMTSMTPPL 66
QY 45 -----NITGYLKLPLK----- 56

DB 67 VSMRLDPLQRYEADLALELLEGERNAQNGHLRYLAERHATFNEARQMWERYNG 126

QY 57 DIIDLKGGIASDLIEIGTSYTHAIPPLPL--SRVEAQVQDRVKKELFELSPKGF 114
DB 127 DLVTAFAKQFQDSNLEITTCGATHGYLPKMKYPOAWAQIYVACEHYEETTCRPPKGIW 186

QY 115 LPELAYDPIIPAILKONGEYELFADGAMLFSAHLSAIPKIPPL----- 159

DB 187 LPECAVYEGLEMLADAGLRYFTDGHGILYARPRFGTYAPIFTGTGVAAFGRDHES 245

QY 160 -----YPLIKAKR-----EKFRYISYLLGLRELKAKLVFEGKVT----- 197

DB 246 SQQVSSQVGPVGPVREFYKDLGWEAEYIYKPYIMPNGQRNKIGIKYH-KITGRGLG 304

QY 198 --LKAVKDIEAVPVVAVVAVTAVMLGIGRLPLMNPKKVASWIEDKNIL--LYGT----- 247

DB 305 LSDKALYD---PYWAKEKAA-----EHAANFMYNRQAEHLGYIMQRPPI 347

QY 248 -----DIEFIGYRDIAG-----YRMSVEGLLEVIDELNSP-----LCLPSELKH 286

DB 348 IVSPYDAELFGHWVEGFWFDYLFKRSWQDQGYAMTHLADYLRNEPTQVCRPSQ--- 404

QY 287 SGRELYLRTSSWAPDKSLRIWREDEGNA-----RLNMLSYNMRELALLAENSARGWEPL 342

DB 405 -----SSWGKYGPFHEYL-NETNAWIYPLHLKAAERMI-EISTL-EPDELGWAL 452

QY 343 PERRLDAPRAIYNDW 357

DB 453 NQAARELLLAQSSDW 467

RESULT 3
S76831
hypothetical protein - *Synechocystis* sp. (strain PCC 6803)
C:Species: *Synechocystis* sp.
A:Variety: PCC 6803

[illegible]

Db 356 HWFEGPDPINAFIRKSAEDWTSEYLIPTTEYLKNNWVQCSSPSP-----SSW 404
QY 358 RGENGE 363
Db 405 -GENDG 409
RESULT 7
E75206
A: alpha-amylase (or 4-alpha-glucanotransferase) PAB0118 - Pyrococcus abyssi (strain Orsay)
C: Species: Pyrococcus abyssi
C: Date: 20-Aug-1999 #sequence_revision 20-Aug-1999 #text_change 20-Jun-2000
C: Accession: E75206
R: anonymous, Genoscope
submitted to the EMBL Data Library, July 1999
A: Description: Pyrococcus abyssi genome sequence: insights into archaeal chromosome structure
A: Reference number: A75001
A: Accession: E75206
A: Status: preliminary
A: Molecule type: DNA
A: Residues: 1-655 <KAW>
A: Cross-references: GB:A248283; GB:AL096836; NID:g5457433; PIDN:CAB49100.1; PID:g545760
A: Experimental source: strain Orsay
C: Genetics:
A: Gene: amyA; PAB0118
C: Superfamily: Dictyoglomus thermophilum amylase A
Query Match 7.8%; Score 145.5; DB 2; Length 655;
Best Local Similarity 22.4%; Pred. No. 0.0046;
Matches 90; Conservative 55; Mismatches 127; Indels 129; Gaps 22;
QY 25 EXAYIPVETLIKKEIP---FGLNITGVLKFLPK---DIIDLKVGSIASDLIEIICTSY 78
Db -28 EXAYRPFLEIL--EYPPNKKVAIHISGILVLEENKPDYIDLKSLVRKQVEIVVAGF 85
QY 79 TTAIPLPLLSRVEAOVRDREKVELFELSPKGFWLPDELAYDPIIPAILKNDGNEYELFA 138
Db 86 YEFVLAAPKEDRLQIYLLKEWAKKI--GYDAKGLWTERVMQPELVKTLREAGIEYVVV 144
QY 139 DGEAMLFSAHLNSA-IKPIKPLYPHLIKAQRE-----KRFYR-----ISYLL 179
Db 145 D-----DYHMSAGSLKDLQFNPPYTDGGEVITVPIDSKLRYLIPFPVDKVISYLH 198
QY 180 GURELRKAIKLVFEGKVTLKAVKDIEAVPVVAVNTAVMIGRLPLMNPKKVASWEDK 239
Db 199 SLASEDSKVAVFH-----DDGEKEGIW-----PM-----TYEWYIEK 231
QY 240 D-----NLLYGTDE-----FIGYRDIAGY-----RMSVEG 266
Db 232 GWLREFFDRVSSDEAINIMLYSEYLOKEPKGLVYLPISAFEMSEWSLPAQQAQLFVE- 290
QY 267 LLEVIDELANSELCLPSELKHSGRELYLTSSWAPDKSLRWDEGN---ARLNKLSYNN 323
Db 291 FVEKLKELN-----MFERYRVFVGGIW---KNP-FYKYPEANYMKRMLNLS--- 334
QY 324 RGEALLAENSARGWEPLRERRDLRAIYND--WRGENG 362
Db 335 -----RLLRDNPSARF-----VLRACNDAYWHGVFG 362
RESULT 8
B64501
A: alpha-amylase (EC 3.2.1.1) - Methanococcus jannaschii
C: Species: Methanococcus jannaschii
C: Date: 13-Sep-1996 #sequence_revision 13-Sep-1996 #text_change 21-Jul-2000
C: Accession: B64501
R: Rult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake,
; Reich, C.I.; Overbeek, R.; Kirkness, E.F.; Weinstein, K.G.; Merrick, J.M.; Glodek, A.;
rson, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A.
Science 273, 1058-1073, 1996
A: Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Woese, C.
A: Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii

A: Reference number: A64300; MUID:96337999
A: Accession: B64501
A: Status: preliminary; nucleic acid sequence not shown; translation not shown
A: Molecule type: DNA
A: Residues: 1-467 <BUL>
A: Cross-references: GB:U67601; GB:L77117; NID:g2826439; PIDN:AAB99631.1; PID:g1592212
C: Genetics:
A: Map position: FOR1586396-1587799
C: Function:
A: Description: catalyzes the hydrolysis of internal 1,4-alpha-D-glucosidic bonds
A: Pathway: glycogen/starch degradation
C: Keywords: glycosidase; hydrolase; polysaccharide degradation
Query Match 7.5%; Score 141.5; DB 2; Length 467;
Best Local Similarity 22.8%; Pred. No. 0.0056;
Matches 86; Conservative 66; Mismatches 130; Indels 95; Gaps 24;
QY 14 EIPKSEIPKVEKAVIPVIEPLIK--BEIPFGLN--ITG-----YTLKFLPKD-IIDLKVG 64
Db 34 KLNKEVFNKVKNCYIPTNEILILEIDYDFKNYSITGVFEQALEF--NDYVLDLFD 91
QY 65 GIASDLIEIIGTSTHAILPL--SRVEAOVRDREKVELFELSPKGFWLPDELAYDPI 123
Db 92 LVKTGNVELIAETVYHSLTSLFETEDFIEDIEMHRKMYXEIFGAKVFRNTELYNNR 151
QY 124 IPAILKNDGVEYLPADG--EAMLFSAHLNSAIKPIKPLYPHLIKAQREKRFYIS---YLL 179
Db 152 IAKIADLGFKAIFTEGIEKIL-----GWSPNLYQSPDGMKI 190
QY 180 GURELRKAIKLVFEGKVTLKAVKDIEAVP-----VWAVNTAVMIGRLPLMNPKKV 232
Db 191 LLRNRLSDDIGR-----FSARDWDQYPLTADRYATWLASTFGEVINI----YMDYETF 241
QY 233 AS--WIEDKDNILLYGTDI-EFIGYR--RMSEVGLLEVIDELNSELCLPSELKH 286
Db 242 GEHMKK-----TGIFELRYLPIEIAKHEHELVNVVSEVVDLEPR----- 283
QY 287 SGRERYLR---TSSWA-PDKSLRWDEGNARLNLMG-SYNNRGELA--LLAENSARGWEP 341
Db 284 --GEIVYHEFATISWADTERDVSALW---GN-KMORISFEKLKDIGKFIKENSKL---- 333
QY 342 LPERLDAFRAIYNDWR 358
Db 334 ---KKLNKFEDELYKMYK 347
RESULT 9
A49512
A: alpha-amylase (EC 3.2.1.1) - Pyrococcus furiosus
C: Species: Pyrococcus furiosus
C: Date: 10-Nov-1995 #sequence_revision 10-Nov-1995 #text_change 21-Jul-2000
C: Accession: A49512
R: Laderman, K.A.; Asada, K.; Uemori, T.; Mukai, H.; Taguchi, Y.; Kato, I.; Anfinsen,
J. Biol. Chem. 268, 24402-24407, 1993
A: Title: alpha-Amylase from the hyperthermophilic archaeobacterium Pyrococcus furiosus
A: Reference number: A49512; MUID:94043280
A: Accession: A49512
A: Status: preliminary
A: Molecule type: DNA
A: Residues: 1-649 <LAD>
A: Cross-references: GB:L22346; NID:g347939; PIDN:AAA72035.1; PID:g347940
C: Genetics:
A: Gene: amya
A: Start codon: GTG
C: Function:
A: Description: catalyzes the hydrolysis of internal 1,4-alpha-D-glucosidic bonds
A: Pathway: glycogen/starch degradation
C: Superfamily: Dictyoglomus thermophilum amylase A
C: Keywords: glycosidase; hydrolase; polysaccharide degradation
Query Match 7.5%; Score 141.5; DB 2; Length 649;

A:Reference number: A75001
A:Accession: A75207
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-1362 <RAW>
A:Cross-references: GB:AJ248283; GB:AL096836; NID:95457433; PIDN:CAB49104.1; PID:el51499
A:Experimental source: strain Orsay
C:Genetics:
A:Gene: apu; PAB0122

Query Match 6.6%; Score 124; DB 2; Length 1362;
Best Local Similarity 21.5%; Pred. No. 0.55;
Matches 82; Conservative 57; Mismatches 142; Indels 100; Gaps 16;

QY 32 IETLIKEEIPFGLNITGVTLPKPKDIIDLKGGTASDLIEITIGTSYTHAILPLP--- 87
DB 236 VETVLKQHM-WLNNHT-----FEEHEKINLLG---NGNVEVTVPVTHPIGPILNDEGW 266
QY 98 LSRVQAQVORDREVKEELF---ELSPKGFWLPPELAYDPIIPAILKNDNGYEVLPADG---E 141
DB 287 YEDFDAQVKKANELYKEYLGAGKVTGKGGAAESALNDKTLLEILAENGKRWMTDQLVLE 346
QY 142 AMLFSAHLNSAIKP-----IKPLYPHLIKA 166
DB 347 KLGVPKTIESYKVPWVAQFGDKKIYLFPRNHDLSRVGFRVAGMNOYDAVKNFVEELKI 406
QY 167 QREKFRYSYLLGLRELKAIKLVFEGKVLKAV-KDIEAVPVVAUNTAV-----ML 219
DB -407 QKQNYDGSIVYVITLDGENPWEHYFDFGKLFLEELYROLELQKKGILRTVTPSEYEMF 466
QY 220 G--IGRLPLMNPKKVASMTEDKDNILLXTDIEFTGYRDIAGYRMSVEGLEVIDELNSE 277
DB 467 GOKANKLTPKMKRLDFTEDNVNALLKAKUL-----GELYDMGVTE----- 509
QY 278 LCLPSELKSGRELVLRTSSWAPDKSLRIWRDEGNARLNLMLSNMRGELALLAENS-D-A 336
DB 510 -----EMQWPESSWT-DGTLSW---IGEPQENIAWYLYLARKALFENKDNV 553
QY 337 RGWEPLPERRDLAFRAIYNDW 357
DB 554 KDNKAYEY---LFRAGSDW 571

RESULT 13
AC1970
hypoetical protein alr1310 [imported] - Anabaena sp. (strain PCC 7120)
A:Note: Anabaena sp. (strain PCC 7120) is a synonym of Nostoc sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 11-Jan-2002
C:Accession: AC1970
R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S. DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena sp. strain PCC 7120
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AC1970
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-744 <GB>
A:Cross-references: GB:BA000019; PIDN:BA073267.1; PID:g17130657; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: alr1310

Query Match 6.6%; Score 123.5; DB 2; Length 744;
Best Local Similarity 22.6%; Pred. No. 0.25;
Matches 79; Conservative 49; Mismatches 111; Indels 111; Gaps 19;

QY 68 SDIEIIGTSYTHAILPL-----LPLSR-----VEAQVORDREVKEELF 106
DB 208 SGOLEVTTPYTHPLPLADNNSGRVAVPNMALPESFQWSEDIPLRLKAWELYTERF 267

QY 107 ELSPKGFWLPPELAYDP-IIPAILKNDNGYEVLPAD-----GEAMLFSAHLNSAIKPIKP--L 159
DB 268 GQEPKGLWPSEQSVSPDILPYIIR-QGFQWICSDAVALGWTLKHFFHRDAGAGNVQOPELL 326
QY 160 Y-PHLIKAQ-----REKFRYSYLLGLRELKAIKLVFEGKVTLPKAVKDIEAVPVW 210
DB 327 YRPYRLATPAGDLAIVFRDHR---LSDLIGF-----TYGAMPKAKQAADL 368
QY 211 VAVNPAVMLGIGRLPLMNPKKVASWIEDKNILLYGTD---IEFIGYRDIAGY-RMSVE 265
DB 369 VCHLQAIKMQRERPERSEQPLVLTALDGENCWEEYPQDGKPFLEAL-YQSUSNEPHIKIV 427
QY 266 GLELVDELNSLCLPSELKSGRELVLRTSSWAPDKSLRIWRDEGNARLNLMLSYNMRG 325
DB 428 TVSEFIEEPATATIPAEQLHSG-----SMW-DGSFTTWIGDP----- 464
QY 326 ELALLAENSADARGWEPLPERL-----DAFRAIY-----NDW 357
DB 465 -----AKN---RANDYLTFAIRMLANHPATEENNPEAWAELYAAGSDW 506

RESULT 14
B69553
methanol dehydrogenase regulatory protein (moxr) homolog - Archaeoglobus fulgidus
C:Species: Archaeoglobus fulgidus
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 10-Sep-1999
C:Accession: B69553
R:Klenk, H.P.; Clayton, R.A.; Tomb, J.F.; White, O.; Nelson, K.E.; Ketchum, K.A.; Dodson, J.; Fleischmann, R.D.; Quackenbush, J.; Lee, N.H.; Sutton, G.G.; Gill, S.; Kirkness, E.; Glodek, A.; Zhou, L.; Overbeek, R.; Gocayne, J.D.; Weidman, J.F.; McDonald, L. Nature 390, 364-370, 1997
A:Authors: Utterback, T.; Cotton, M.D.; Spriggs, T.; Artlach, P.; Kaine, B.P.; Sykes, Smith, H.O.; Woese, C.R.; Venter, J.C.
A:Title: The complete genome sequence of the hyperthermophilic, sulfate-reducing archaean, *Archaeoglobus fulgidus*.
A:Reference number: A69250; MUID:98049343
A:Accession: B69553
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-324 <KLE>
A:Cross-references: GB:AB001108; GB:AB000782; NID:G2689431; PIDN:AAB91247.1; PID:g265
C:Superfamily: methanol dehydrogenase regulatory protein

Query Match 6.4%; Score 120.5; DB 1; Length 324;
Best Local Similarity 21.7%; Pred. No. 0.13;
Matches 71; Conservative 60; Mismatches 121; Indels 75; Gaps 14;

QY 1 LRALVFHGNLQVAEIP---KSEIPKVIKAVIPVITETLIKKEIPEGLNITGYTLKFLPKD 57
DB 35 LAAALTNGLIFEDYFGLGKTLILAKVAVR-----IGADYRVQFTDPLLPD 82
QY 58 IIDLVKGGI---ASDLIEII-GTSTHAIL-----PLPLSRVQAQVORDREVKEEL 105
DB 83 II-----GVKIRGDRFEPVKGPIFTNVLLADEINRSPKPTQAALLEAMEEKQITVEGET 137
QY 106 FELSPKGFWLPPELAYDPIIPAILKNDNGYEVLPADGEAMLFSAHLNSAIKPIKPLYPHILK 165
DB 138 FLSLMPFFVL--ATQNPI-----EQETYPLEAQMDRFLMRPQYPSIE 182
QY 166 AQREKFRYSY-----LIGRELKRAIKLVFEGKVTLPKAVKDIEAVPVW 211
DB 183 EEMEILRRRISWRKDDPTEDVEPVYSLETFRRIQDAVEAVYVDSILKYISELVRA---T 239
QY 212 AVNTAVNLGI---GRPLMNPKKVASWIEDKNIL---LYGTDIEFIGYRDIAGYRMSVE 265
DB 240 REHELVELSSPRGGALLKLARAVMDGRDFVTPDDVKRVAVEALAHRVILAFAYE 299
QY 266 GLL--EVIDELNSLCLPSELKSGRE 290
DB 300 GLRAEEVVEILNSVRP---KYEAQE 323

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 2, 2002, 18:05:58 ; Search time 29.22 Seconds
(without alignments)
2155.036 Million cell updates/sec

Title: US-09-886-400-4
Perfect score: 1877
Sequence: 1 LRLVFNGLQYAEIPKSEI.....RRDAFRALYNDRWGENGEP 364

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_l9:*

- 1: sp_archaea.*
- 2: sp_bacteria.*
- 3: sp_fungi.*
- 4: sp_human.*
- 5: sp_invertebrate.*
- 6: sp_mammal.*
- 7: sp_mhc.*
- 8: sp_organelle.*
- 9: sp_phage.*
- 10: sp_plant.*
- 11: sp_rodent.*
- 12: sp_virus.*
- 13: sp_vertebrate.*
- 14: sp_unclassified.*
- 15: sp_rvirus.*
- 16: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1532	81.6	364	1 Q9HHB5	Q9hhb5 pyrococcus
2	1494	79.6	364	17 O58106	O58106 pyrococcus
3	162.5	8.7	529	16 P74630	P74630 synchocyst
4	149.5	8.0	378	17 Q97BM4	Q97bm4 thermoplasm
5	147	7.8	527	16 Q97GF3	Q97gf3 clostridium
6	142.5	7.6	443	17 Q973T0	Q973t0 sulfolobus
7	136.5	7.3	895	17 Q972N0	Q972n0 sulfolobus
8	130.5	7.0	560	17 Q50094	Q50094 pyrococcus
9	128	6.8	923	16 Q9KD04	Q9kd04 bacillus ha
10	124	6.6	1362	17 Q9V294	Q9v294 pyrococcus
11	122.5	6.5	357	17 Q9HL91	Q9hl91 thermoplasm
12	120.5	6.4	324	17 Q30246	Q30246 archaeoglob
13	119	6.3	602	17 Q9V0M7	Q9v0m7 pyrococcus
14	111	5.9	526	16 O83377	O83377 treponema p
15	109.5	5.8	447	17 Q972D2	Q972d2 sulfolobus
16	106	5.6	526	16 O53278	O53278 mycobacteri

17	105.5	5.6	902	17	Q97YY0	Q97yy0 sulfolobus
18	104	5.5	704	10	Q9ZVN2	Q9zvn2 arabidopsis
19	103.5	5.5	458	10	Q9MBD0	Q9mbd0 pyrus pyrif
20	103.5	5.5	701	2	Q93MG7	Q93mg7 thiobacillu
21	103	5.5	754	13	O13131	O13131 oncorhynch
22	102	5.4	754	13	O13132	O13132 oncorhynch
23	100.5	5.4	266	16	Q97M42	Q97m42 clostridium
24	100.5	5.4	619	16	O83182	O83182 treponema p
25	99	5.3	4845	11	O88738	O88738 mus musculu
26	98	5.2	503	5	O17585	O17585 caenorhabdi
27	98	5.2	764	15	O67347	O67347 aquifex aeo
28	97.5	5.2	402	2	O68779	O68779 versinia pe
29	97.5	5.2	605	16	Q9ZDR2	Q9zdr2 rickettsia
30	97.5	5.2	627	10	Q9ZQF1	Q9zqf1 arabidopsis
31	97.5	5.2	684	10	Q9M2S4	Q9m2s4 arabidopsis
32	97.5	5.2	888	16	Q98PQ2	Q98pq2 mycoplasma
33	97.5	5.2	1085	15	O25577	O25577 helicobacte
34	97	5.2	510	16	O66973	O66973 aquifex aeo
35	97	5.2	531	17	O97BF8	O97bf8 thermoplasm
36	97	5.2	610	2	O32582	O32582 escherichia
37	97	5.2	929	5	Q9VR32	Q9vr32 drosophila
38	96.5	5.1	197	13	Q918E8	Q918e8 fugu rubrip
39	96	5.1	636	2	O87139	O87139 vibrio chol
40	96	5.1	636	2	O34235	O34235 vibrio chol
41	96	5.1	1150	16	Q99UY8	Q99uy8 staphylococ
42	95.5	5.1	1134	17	Q9UV58	Q9uv58 pyrococcus
43	95.5	5.1	1326	5	O16928	O16928 caenorhabdi
44	95	5.1	296	2	Q9Z597	Q9z597 streptomyce
45	95	5.1	394	17	Q9UYC0	Q9uyc0 pyrococcus

ALIGNMENTS

RESULT	ID	PRELIMINARY	PRT	364 AA
Q9HHB5	Q9HHB5			
AC	Q9HHB5			
DT	01-WAR-2001 (Tremblrel. 16, Created)			
DT	01-WAR-2001 (Tremblrel. 16, Last sequence update)			
DT	01-DEC-2001 (Tremblrel. 19, Last annotation update)			
DE	ALPHA-GALACTOSIDASE			
GN	GALA			
OS	Pyrococcus furiosus			
OC	Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus			
OX	NCBI_TaxID=2261			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=DSM 3638			
RA	Verhees C.H.			
RL	Submitted (OCT-1999) to the EMBL/GenBank/DBJ databases			
DR	EMBL; AF195244; AAG28455.1			
DR	InterPro; IPR004300; Glyco_hydro.57			
DR	Pfam; PF03065; Glyco_hydro.57; 1			
SQ	SEQUENCE 364 AA; 41545 MW; 3E8E0AF5BDCCF2A5 CRC64;			

Query Match 81.6%; Score 1532; DB 1; Length 364;
Best Local Similarity 79.1%; Pred. No. 4.9e-117;
Matches 287; Conservative 38; Mismatches 38; Indels 0; Gaps 0;

Qy	1	LRLVFNGLQYAEIPKSEIPKVEKAYIPVETLIKKEIPFGLNITGYTLKFLPKDIID	60
Db	1	MRALVFNGLQYAEIPKSEIPKVEKAYPTTSELRIPFGLNITGYSLFLPKDLIA	60
Qy	61	LVRGGIASDLIEIGTSYTHAILPLLSRVEAQVORREVEKEELFSPKGFWLPCLAY	120
Db	61	LRKEGIESGLIEIGTSYTHAILPLLSRVEAQIKRDREVKENILEVSPGFWLPCLAY	120
Qy	121	DPPIPAIKDNGVEYLFADGEAMLFSAHLNSAIKPKPLYPHLIKAQREKFRYSYLLG	180
Db	121	DPPIPAIKDNNVEYLFADGEAMLFNSHLNSAIKPKPLYPHLIKAQRGGLVYLLG	180

```

QY 181 LRELKAIKLVFEKGVTKAVKDIEAVPVVAVNTAVMLGIGRLPLMNPKNKVASWIEDKD 240
DB 181 LRELKAINLVFEKGVTKAVKDIEAVPVVAVNTAVMLGIGRLPLMNPKNKVASWIEDKD 240
QY 241 NILLYGTDIEFGYRDIAGYRMSVGEGLVEIDELNSCLPSELKHSGRGLYLRSSWAP 300
DB 241 EILLYGTDIEFGYRDIAGYRMSVGEGLVEIDELNSCLPSELKHSGRGLYLRSSWAP 300
QY 301 DKSIRIWEDEGNARLNMLSYNMRGELALLAENSARGWEPLPERRLDFAFRALYNDRGE 360
DB 301 DKSIRIWEDEGNARLNMLSYNMRGELALLAENSARGWEPLPERRLDFAFRALYNDRGE 360
QY 361 NGE 363
DB 361 NGK 363

RESULT 2
O58106 PRELIMINARY; PRT; 364 AA.
AC O58106;
DT 01-AUG-1998 (TrEMBLrel. 07, Created)
DT 01-AUG-1998 (TrEMBLrel. 07, Last sequence update)
DE 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE HYPOTHETICAL 41.8 KDA PROTEIN PHC368.
GN PHC368.
OS Pyrococcus horikoshii.
OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
OX NCBI_TaxID=53955;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=98344137; PubMed=9679194;
RA Kavaravayasi Y., Sawada M., Horikawa H., Haikawa Y., Hino Y.,
RA Yamamoto S., Sekine M., Baba S.-I., Kosugi H., Hsuo-Yana A., Nagai Y.,
RA Sakai M., Ogura K., Otsuka R., Nakazawa H., Takamiya M., Ohfuku Y.,
RA Funahashi T., Tanaka T., Kudoh Y., Yamazaki J., Kushida N., Oguchi A.,
RA Aoki K.-I., Foshizawa T., Nakamura Y., Radd F.T., Horikoshi K.,
RA Masuchi Y., Shizuya H., Kikuchi H.;
RT "Complete sequence and gene organization of the genome of a hyper-
RT thermophilic archaeobacterium, Pyrococcus horikoshii OT3.";
RL DNA Res. 5:55-76(1998).
DR EMBL: AP000002; BAA29442.1;
DR InterPro: IPR004300; Glyco_hydro_57.
DR Pfam: PF03065; Glyco_hydro_57; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 364 AA; 41755 MW; 7B4B36AB4A975BAD CRC64;

```

```

Query Match 79.6%; Score 1494; DB 17; Length 364;
Best Local Similarity 75.8%; Pred. No. 6.3e-114;
Matches 275; Conservative 49; Mismatches 39; Indels 0; Gaps 0;

QY 1 LRALVHGNLQYAEIPKSEIPKVIKAYIPVETLKEIPFGLNITGYTLKPLPKDIID 60
DB 1 MRALIFHGNLQYAEIPKSEIPKVIKAYIPVETLKEIPFGLNITGYTLKPLPKDIID 60
QY 61 LVKGGIASDLIEIGTSYTHAILPLPL--SRVEAQVQDRREVKEELFELSPKGF 120
DB 61 LKKEGISELIEIGTSYTHAILPLPL--SRVEAQVQDRREVKEELFELSPKGF 120
QY 121 DPTIIPAILKNDGYEYLFADGEMALFSAHLNSAIKPIKPLYPHLIKAKRFRFYISYLLG 180
DB 121 DPTIIPAILKNDGYEYLFADGEMALFSAHLNSAIKPIKPLYPHLIKAKRFRFYISYLLG 180
QY 181 LRELKAIKLVFEKGVTKAVKDIEAVPVVAVNTAVMLGIGRLPLMNPKNKVASWIEDKD 240
DB 181 LRELKAINLVFEKGVTKAVKDIEAVPVVAVNTAVMLGIGRLPLMNPKNKVASWIEDKD 240
QY 241 NILLYGTDIEFGYRDIAGYRMSVGEGLVEIDELNSCLPSELKHSGRGLYLRSSWAP 300
DB 241 EILLYGTDIEFGYRDIAGYRMSVGEGLVEIDELNSCLPSELKHSGRGLYLRSSWAP 300

```

```

QY 301 DKSIRIWEDEGNARLNMLSYNMRGELALLAENSARGWEPLPERRLDFAFRALYNDRGE 360
DB 301 DKSIRIWEDEGNARLNMLSYNMRGELALLAENSARGWEPLPERRLDFAFRALYNDRGE 360
QY 361 NGE 363
DB 361 NGK 363

RESULT 3
P74630 PRELIMINARY; PRT; 529 AA.
AC P74630;
DT 01-FEB-1997 (TrEMBLrel. 02, Created)
DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE HYPOTHETICAL 62.1 KDA PROTEIN.
GN SLL0735.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX NCBI_TaxID=1148;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97061201; PubMed=8905231;
RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyajima N., Hiroseawa M., Sugiyama M., Sasamoto S., Kimura T.,
RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naruo K., Okumura S.,
RA Simpo S., Takeuchi C., Wada T., Watanabe A., Yamada M., Yasuda M.,
RA Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions.";
RL DNA Res. 3:109-136(1996).
DR EMBL: D90916; BAA:8743.1;
DR InterPro: IPR003803; DUF200.
DR Pfam: PF02651; DUF200; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 529 AA; 62075 MW; 4C1A5048A784E30 CRC64;

```

```

Query Match 8.7%; Score 162.5; DB 16; Length 529;
Best Local Similarity 21.4%; Pred. No. 7.1e-05;
Matches 101; Conservative 47; Mismatches 140; Indels 183; Gaps 19;

QY 3 ALVPHGNLQYAEIPKSEIPKVIKAYIPVETLKEIPFGLNITGYTLKPLPKDIID 47
DB 7 ALVPHGNLQYAEIPKSEIPKVIKAYIPVETLKEIPFGLNITGYTLKPLPKDIID 47
QY 48 -----GYTLKPLPK----- 56
DB 67 VSMRLDPLQORYEAHLSLLQELLAKAIVRNEHNGHLQYLAIFYAKEFAAIRETWERYDG 126
QY 57 DIIDLKGGIASDLIEIGTSYTHAILPLPL--SRVEAQVQDRREVKEELFELSPKGF 114
DB 127 DLVTAFOQFQDSNNLEITGCGATHGYFPLMKYPOAVWQAIKVACEHYEENFGSRSPK 186
QY 115 LPELAYDPIIPAILKNDGYEYLFADGEMALFSA-----HLNS 151
DB 187 LPELAYDPIIPAILKNDGYEYLFADGEMALFSA-----HLNS 151
QY 152 -----AIKPIKPLYPHLIK-AQREKFRFYISYLLGLRELKAIKLVF-----BGKVT 197
DB 247 QQVWSSQVGYGPDVYREFYKDLGWEAEYIKRYPMPNGQRNIGIKYHKITSRDGGLS 306
QY 198 LKAVKDIKAVPVVAVNTAV-----MLGI-GRPLMNPKNKVAS-----WI 236
DB 307 EKAVD-----PYWAKAAEAASNFNMQQQVGHLSGNGRPLVVSYPDAELFGHWY 362
QY 237 E-----DKDNI-----LLYGTDIETI-----GYRDIAGY----- 260
DB 363 EGPWFIDYLFKRSWFDQDTFMTHLADYLRGNPHQVQCRPSQSSWGKGFHYEYWLNDTNA 422
QY 261 -----RMSVEGLLEVIDELNSCLPSELKHSGRGLYLRSSWAP 299

```

Db 423 WYPHLUHKAERMIELSHREAVDELEK-----ALNQAARELLLAQSSDWA 468

RESULT 4
Q97BM4 PRELIMINARY; PRT; 378 AA.
ID Q97BM4;
AC Q97BM4;
DT 01-OCT-2001 (TReMBLrel. 18, Created)
DT 01-OCT-2001 (TReMBLrel. 18, Last sequence update)
DT 01-OCT-2001 (TReMBLrel. 18, Last annotation update)
DE ALPHA-AMYLASE.
GN TVG0421416.
OS Thermoplasma volcanium.
OC Archaea; Euryarchaeota; Thermoplasmatales; Thermoplasmaceae;
OC Thermoplasma.
OX NCBI_TaxID=50339; *
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=GSSI / DSM 4299 / JCM 9571;
RX MEDLINE=20570456; PubMed=11121031;
RA Kawashima T., Amano N., Koike H., Makino S.-I., Higuchi S.,
RA Kawashima-Chiya Y., Watanabe X., Yamazaki M., Kanehori K., Kawamoto T.,
RA Nunohiba T., Yamamoto Y., Atamaki H., Makino K., Suzuki M.:
"Archaeal adaptation to higher temperatures revealed by genomic
RT sequence of Thermoplasma volcanium".
RL Proc. Natl. Acad. Sci. U.S.A. 97:14257-14262(2000).
DR EMBL; AP000992; BABS9573.1.; -.
KW Complete proteome.
SQ SEQUENCE 378 AA; 44957 MW; 49FCCF63BE6B3A7 CRC64;

Query Match 8.0%; Score 149.5; DB 17; Length 378;
Best Local Similarity 21.1%; Pred No. 0.00051;
Matches 66; Conservative 49; Mismatches 123; Indels 75; Gaps 11;

QY 22 KVIERAYIPVIETLKESIPFLGNITG-----YTLKFLPKDIIDLKYGSIADLIEI 73
Db 42 :|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 42 RVAERSYIPATNLMEYGIKSFSITAVEQALMYNTK----VDAIDDYVKSGLCGM 96
QY 74 IGSYTHALPLLELSRYEAOQRDERVKELFELSPKGFWLPDELAYDPILPAIKLDNGY 133
Db 97 LSEITYHSLSWNYDFEKQVDMDRLMKRFNVVPKVFNTELYDDRIAEMVKRWGF 156
QY 134 EYLFADGEAMLFSAHNSAIKPILPFLPHLIKAORERFRYS---YLGLRELRAIKL 190.
Db 157 TSIIPEG-----TDSIVKDHPNV-----RYASPSGLNLYLRANYMSDN 196
QY 191 VFGKGYTLKAVRDIEAPVWAVNTAVMLGICRPLMPKKVASWTEDKNILLTGTDIE 250
Db 197 SPRESNT--KKRDYP-----LTADKYAKWINESEGDMVN---L 229
QY 251 FTGYRDIAGYRMSVEGLEVIDEL-----NSECLPSELKHGRELYL--RTSWA 299
Db 230 FMDYTFGHQTQGTGFIDFMKILPVFYROYGTIITISEAKRHVRKVDSIPETISWA 289
QY 300 PD-KSLRIWREDE 311
Db 290 DTRRDLRAWLENE 302

RESULT 5
Q97GF3 PRELIMINARY; PRT; 527 AA.
ID Q97GF3;
AC Q97GF3;
DT 01-OCT-2001 (TReMBLrel. 18, Created)
DT 01-OCT-2001 (TReMBLrel. 18, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE UNCHARACTERIZED CONSERVED PROTEIN, RELATED TO
DE ALPHA-AMYLASE/ALPHA-MANNOSIDASE.
GN CAC2414.
OS Clostridium acetobutylicum.
OC Bacteria; Firmicutes; Bacillus; Clostridium group; Clostridiaceae;

QY 221 IGRPLMNPVK--VASWIEDK-----DNLLYGTDEIEI 252
 Db 289 YHRTGHTTEESKDLVREMAEKRVQEHANHTIGAIHETIDQHGQNPFPYVMVTPFDALF 348
 QY 253 GYRDIAGYRMSVEGLEV-----IDELNSELCLPSELKHSRELYLRTSSWAPDKSLRIWR 308
 Db 349 GHWFEQPEW-IEALYEGGADRVSFITPELVORHYQ-DFOTAHVSFSTWGRDGYGHVWL 406
 QY 309 EDEGNARLMSYNWRELALL-----AENSARQWEPLPERLDFAFAYN 355
 Db 407 NDH-NAMRYRHYHREKDLAKIVAMYPOTVLEKQAIQWVREW-----MLAVSS 455
 QY 356 DW 357
 Db 456 DW 457

RESULT 10
 Q9V294 PRELIMINARY; PRT; 1362 AA.
 AC Q9V294;
 DT 01-MAY-2000 (TREMBlrel. 13, Created)
 DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
 DT 01-DEC-2001 (TREMBlrel. 19, Last annotation update)
 DE AMYLOPULLULANASE.
 GN APU OR PAS0122.
 OS Pyrococcus abyssi.
 OC Archaea; Euryarchaeota; Thermococcales; Thermococcaceae; Pyrococcus.
 OX NCBI_TaxID=29292;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ORSAY;
 RA Heilig R.;
 RT "Pyrococcus abyssi genome sequence: insights into archaeal chromosome structure and evolution."
 RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AJ248283; CAB49104.1;
 DR InterPro: IPR004300; Glyco_hydro.57.
 DR Pfam: PF03065; Glyco_hydro.57; 1.
 KW Complete proteome.
 SQ SEQUENCE 1362 AA; 154578 MW; 277AFAB4E14860D1 CRC64;

Query Match 6.6%; Score 124; DB 17; Length 1362;
 Best Local Similarity 21.5%; Pred. No. 0.39;
 Matches 82; Conservative 57; Mismatches 142; Indels 100; Gaps 16;

QY 32 IETLKEEIPGLNTGYTLKPLPKDIDLVKGGIASDLIEIICTSYTHAILPLP--- 87
 Db 236 VETVLKHQW-WLLNHT-----FEEHEKINLLG---NGNVEVTVVPTHTPIGLNDFGW 286
 QY 88 LSRVGAQVORDREVKEELF---ELSPKGFPLPELAYDPIIPAILKDNQGYEYFADG---E 141
 Db 287 YEDFPAQVAKANELYKEVLGACKVTPKGVAAESALNDKTLLEILAENGKWKWTDQLVLE 346
 QY 142 AMLFSAHLNSAIKP-----IKPLYPHILKA 166
 Db 347 KLGYPKTIESYIKPVVAQFGDKKIYLFPRNHDLSDRVGFYAGMNQIDAVKNFYEELKI 406
 QY 167 QREKPRYSYLLGLRELKALKVFECKVTLKAV-KDIEAPVWVAVNTAV-----ML 219
 Db 407 OKONVDGSLVYVITLDGENPWEHYPDFDKLFEELYRLEELQKKGLIRTPSEYIEMF 466
 QY 220 G--IGRLPLMNPVKVASWIEDKDNLLYGTDEIEFGYRDIAGYRMSVEGLEVIDELNSE 277
 Db 467 GUKANKLTPKMKRDLFTEDNVNALLKAKTL-----GELYDMVGVE----- 509
 QY 278 LCLPSELKHSRELYLRTSSWAPDKSLRIWREDEGNARLMSYNWRELALIAENS-D-A 336
 Db 510 -----EMQWPESSWI-DGTLSTW---IGEPQENIAWLYLARKALFENKDNV 553
 QY 337 RGWEPLPERLDFAFAYNDW 357

QY 14 ETPKSEIPKVIKAYIPVIEL-----KKEIPFGLNTGYTLKPLPKDITDL---VK 63
 Db 71 EYIKREFEYMERKLKMSDEDLRFKDEKLREAINF---MIGYF-----KDYVSYWKSID 122
 QY 64 GGIASDL-----IEIGSYTHAILPLPLSR-VEAQQVORDREVKEELFELSFKGEW 114
 Db 123 GNILKFPLEODEGVVEVITSAATHGYLPLGRDEATEAALLNGIKYKFKRGRGIW 182
 QY 115 LPELAYDFI-----IPAILKDNQGYEYLFADGEMLFSAHLNSAIKPKP 158
 Db 183 LPECAIRPDGLWKSPTGEVKKWRKGIEHFLKFGIEYFFVE-----SHLIDK-GPVS 234
 QY 159 LYPHILKAQREKRRFYISYLLGLRELKALKVFECKVTLKAVKDIEAVPVWA 212
 Db 235 RYGNILPA-KIKRSTLRPYF-----LKNGLIAVFARNRET-----GIQWNSA 274

RESULT 9
 Q9KD04 PRELIMINARY; PRT; 923 AA.
 AC Q9KD04;
 DT 01-OCT-2000 (TREMBlrel. 15, Created)
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE BH1415 PROTEIN.
 GN BH1415.
 OS Bacillus halodurans.
 OC Bacteria; Firmicutes; Bacillus/Clostridium group;
 OC Bacillus/Staphylococcus group; Bacillus.
 OX NCBI_TaxID=98665;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C-125 / JCM 9153;
 RX MEDLINE=20512582; PubMed=11058132;
 RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
 RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhara S.,
 RA Horikoshi K.;
 RT "Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and genomic sequence comparison with Bacillus subtilis."
 RL Nucleic Acids Res. 28:4317-4331(2000).
 DR EMBL: AP001512; BAB05134.1;
 DR InterPro: IPR003803; DUF200.
 DR InterPro: IPR001296; Glycos_transf_1.
 DR Pfam: PF02651; DUF200; 1.
 DR Pfam: PF00534; Glycos_transf_1; 1.
 KW Complete proteome.
 SQ SEQUENCE 923 AA; 106638 MW; 8C7B7D64F6303553 CRC64;

Query Match 6.8%; Score 128; DB 16; Length 923;
 Best Local Similarity 20.1%; Pred. No. 0.11;
 Matches 97; Conservative 60; Mismatches 167; Indels 158; Gaps 24;

QY 3 ALVFHGNQYAEIPKSE-----IPKVIKAYIPVIELIKEEIPFGLNI----- 46
 Db 7 SLVLHAHLPVVRHQEDRLEERLWEAMSETYIPLLWALEKLPVKHATISFTPPVMEML 66
 QY 47 -----TGY-----TLKPLPKD-----IIDLVKG-----G 65
 Db 67 SDPLVQTRVNLHNTPEQLKKEERKTNDQNTQNLVQYKRYEKLKATFLQWRNLLIG 126
 QY 66 IASDLIE-----IIGTSYTHAILPLPLSR-VEAQQVORDREVKEELFELSFKGEW 119
 Db 127 FRS-LMENSQCLMTSAATHAFPPVLPKTKAEIRAQVRHGIACFEQHFQKGLPGLWPECA 185
 QY 120 YDPIIPAILKDNQGYEYLFADGAMLF---SAHLNSAIKPKLY-PH---LIKAQREKRF 172
 Db 186 FSPGVDRILLFEIGIRYTFYDEGAVLTADTPHKGS-----APIVSPHGIALFPRHTELSA 241
 QY 173 RYTSYLLGLRELKALKVFECKVTLKA-VKDIEAVPVK-----VAVNTAVMLG 220
 Db 242 KWNSSILG-----YPGDVYREFYRIADRENDYIKPHVHKDGRIDTG---LK 288

Db 554 KWNKAYEY---LFRAEGSDW 571

RESULT 11

Q9HL91 PRELIMINARY; PRT; 357 AA.

ID Q9HL91

AC Q9HL91

DT 01-MAR-2001 (TrEMBLrel. 16, Created)

DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)

DT 01-MAR-2001 (TrEMBLrel. 16, Last annotation update)

DE ALPHA-AMYLASE RELATED PROTEIN.

GN TA0339.

OS Thermoplasma acidophilum.

OC Archaea; Euryarchaeota; Thermoplasmatales; Thermoplasmataceae;

OC Thermoplasma.

OX NCBI_TaxID=2303;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=DSM 1728;

RX MEDLINE=20479972; PubMed=11029001;

RA Ruepp A., Granl W., Santos-Martinez M.-L., Koretke K.K., Volker C.,

RA Mewes H.-W., Frishman D., Stocker S., Lupas A.N., Baumeister W.,

RT "The genome sequence of the thermophilic scavenger Thermoplasma acidophilum";

RL Nature 407:508-513(2000).

DR EMBL; AL445064; CAC11483.1; -

KW Complete proteome.

SQ SEQUENCE 357 AA; 41845 MW; B8C1A97F4EB237F4 CRC64;

Query Match 6.5%; Score 122.5; DB 17; Length 357;

Best Local Similarity 16.6%; Pred. No. 0.075;

Matches 51; Conservative 59; Mismatches 126; Indels 71; Gaps 8;

QY 22 KVEIKAVPIETLIKKEPFGNITGYL-----KFLPKDIIDLVKGGIADLIEIIGTS 77

Db 20 RIANNYPATQSLMDYGRSELSGTMEQAKYCPK-VIDVDIVDSRQCELLSET 78

QY 78 YTHAILDLPLSRVEAQVDRREVKEELFELSPKGFVLPDLPPIIPAILKONGYEYLF 137

Db 79 YHSLASINWDEEFVQRVQMOESAIAKKTNYEFVSFRNTELIYNDHIAEVAKMGFRNII 138

QY 138 ADGEAMLFSAHLSAIPKIPLYPLHKAQRKRPRY-----ISVLLGLRELKRAIKLVF 192

Db 139 AEGTDDIASRY-----DVNRYAAPSGINILYLRNPLSDDISRF 178

QY 193 EGVTLKAVKDIKAVPVAVNFMVAVMLGIGRPLMNPVKVSVASWIEDKNILLYGTDFI 252

Db 179 SNRAWADYPLTADKFAFWISASSGEIVNL-----FM 209

QY 253 GYRDIAGYRMSVGLLEVI-----DELNSLCLPSELKHSREL--YLRTSSWA-P 300

Db 210 DYTFGEHQRPETGIFELRYLPMYFEENDVHTILVREAEARHRTDFISVSKTTSWADK 269

QY 301 DKSLRIW 307

Db 270 NRDLASW 276

RESULT 12

C30246 PRELIMINARY; PRT; 324 AA.

ID C30246

AC C30246

DT 01-JAN-1998 (TrEMBLrel. 05, Created)

DT 01-JAN-1998 (TrEMBLrel. 05, Last sequence update)

DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)

DE METHANOL DEHYDROGENASE REGULATORY PROTEIN (MOXR).

GN AF2425.

OS Archaeoglobus fulgidus.

OC Archaea; Euryarchaeota; Archaeoglobales; Archaeoglobataceae;

OC Archaeoglobus.

OX NCBI_TaxID=2234;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=VC-16 / DSM 4304 / ATCC 49558;

RX MEDLINE=98049343; PubMed=9389475;

RA Klenk H.-P., Clayton R.A., Tomb J.-F., White O., Nelson K.E.,

RA Ketchum K.A., Dodson R.J., Gwinn M., Hickey E.K., Peterson J.D.,

RA Richardson D.L., Kerlavage A.R., Graham M.H., Kyripides N.C.,

RA Fleischmann R.D., Quackenbush J., Lee N.H., Sutton G.G., Gill S.,

RA Kirkness E.F., Dougherty B.A., McKenney K., Adams M.D., Loftus B.,

RA Peterson S., Reich C.I., McNeil L.K., Badger J.H., Glodek A., Zhou L.,

RA Overbeek R., Gocayne J.D., Weidman J.F., McDonald L., Uterback T.,

RA Cotton M.D., Spriggs T., Artiach P., Kaine B.P., Sykes S.M.,

RA Sadow P.W., D'Andrea K.P., Bowman C., Fujii C., Garland S.A.,

RA Mason T.M., Olsen G.J., Fraser C.M., Smith H.O., Woese C.R.,

RA Venter J.C.;

RT "The complete genome sequence of the hyperthermophilic, sulphate-

RT reducing archaeon Archaeoglobus fulgidus";

RL Nature 390:364-370(1997).

DR EMBL; AF001108; AAB91247.1; -

DR TIGR; AF2425; -

DR InterPro; IPR000523; Mg_chelatase_ChII.

DR Pfam; PF01078; Mg_chelatase; 1.

KW Hypothetical protein; Complete proteome.

SQ SEQUENCE 324 AA; 36609 MW; B0F64965EFC82F1F CRC64;

Query Match 6.4%; Score 120.5; DB 17; Length 324;

Best Local Similarity 21.7%; Pred. No. 0.095;

Matches 71; Conservative 60; Mismatches 121; Indels 75; Gaps 14;

QY 1 LRALVPHGNLQVAEIP---KSEIPKVIKAYIPVITELIKEIPPLNITGYTLKFLPKD 57

Db 35 LAALINGNILEDPGLCKTLLAKVFARV-----IGADYRRVQFTDLPSPD 82

QY 58 IIDLVKGI---ASDLIEII-GTSYTHAIL-----PLPLSRVEAQVDRREVKEEL 105

Db 83 II-----GVKINRGDRFEVKGPIETNVLNLADEINRSPKTKAALLEAMEEKQITVEGT 137

QY 106 FELSPKGFVLPDLPPIIPAILKONGYEYLFADGEAMLFSAHLSAIPKIPLYPLHIIK 165

Db 138 FSUSMPFFVL--ATQNP-----EQGTYPDLPDPAQMDRFLMRPQGPESIE 182

QY 166 AQREKFRYISY-----LLGLRELKRAIKLVPEKGVTLKAVKDIKAVPVVW 211

Db 183 EEMEILRRIRISMKKDPTEDVFPVSLTFPRIQDAVAVYVDKSIKYLSELVRA---T 239

QY 212 AVNTAVMLGI---GRPLMNPVKVSVASWIEDKNIL--LYGTDIEFGYRDIAGYRMSVE 265

Db 240 REHELVELGSSPRGGALLKLARALAVMDGRDFVDPDDVKRVAVELAHRLVILFEYAVE 299

QY 266 GLL--EVIDELNSLCLPSELKHSRE 290

Db 300 GLRAEEVVEILNSVRVP---KYEAGE 323

RESULT 13

Q9VOM7 PRELIMINARY; PRT; 602 AA.

ID Q9VOM7

AC Q9VOM7

DT 01-MAY-2000 (TrEMBLrel. 13, Created)

DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)

DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)

DE HYPOTHETICAL 70.2 KDA PROTEIN.

GN PAB1857.

OS Pyrococcus abyssi.

OC Archaea; Euryarchaeota; Thermococcales; Thermococaceae; Pyrococcus.

OX NCBI_TaxID=29292;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=ORSAY;

RA Helling R.;

RT "Pyrococcus abyssi genome sequence: insights into archaeal chromosome structure and evolution";

RL Submitted (JUL-1999) to the EMBL/GenBank/DBJ databases.

EMBL: AJ248285; CAB49676.1; -.
InterPro: IPR001395; Aldoc_ket-red.
InterPro: IPR003803; DUF200.
InterPro: IPR003583; HHH_1.
Pfam: PF02651; DUF200; 1.
SMART: SM00378; HHH1_1.
PROSITE: PS00062; ALDO-KETO-REDUCTASE_2; UNKNOWN_1.
Hypothetical protein: Complete proteome.
SEQUENCE 602 AA; 70219 MW; 346B1198D54D43E CRC64;
SEQUENCE

[illegible]

RESULT	14
ID	083377
AC	PRELIMINARY;
DC	PRT; 526 AA.
DT	083377;
DT	01-NOV-1998 (TREMBLrel. 08, Created)
DT	01-NOV-1998 (TREMBLrel. 08, Last sequence update)
DT	01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE	CONSERVED HYPOTHETICAL PROTEIN.
GN	TP0358.
OS	Treponema pallidum.
OC	Bacteria; Spirochaetales; Spirochaetaceae; Treponema.
OX	NCBI_TaxID=160;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=NICHOLS;
RX	MEDLINE=98332770; PubMed=9665876;
RA	Fraser C.M., Norris S.J., Weinstock G.M., White O., Sutton G.G.,
RA	Dodson R., Guinn M., Hickey E.K., Clayton R., Ketchum K.A.,
RA	Sodergren E., Hardham J.W., McLeod M.P., Salzberg S., Peterson J.,
RA	Khalat H., Richardson D., Howell J.K., Chidambaram M., Utterback T.,
RA	McDonald L., Artlich P., Bowman C., Cotton M.D., Fujii C., Garland S.,
RA	Vatsh B., Horst K., Roberts K., Sandusky M., Weidman J., Smith H.O.,
RA	Hunter J.C.;
RT	"Complete genome sequence of Treponema pallidum, the syphilis
RT	spirochete.";
RL	Science 281:375-388(1998).
DR	EMBL; AE001215; AAC65344.1; -.
DR	TIGR; TP0358; -.
DR	InterPro; IPR003803; DUF200.
DR	Pfam; PF02651; DUF200; 1.
SK	Complete proteome.
QW	SEQUENCE 526 AA; 59699 MW; D5F0C32CADB6C4IA CRC64;

[illegible]

```

QY 65 GIASDL-----TEITGSTYTHAILPL---LPLSRVDAQQRDRVYKEELFELSPKGF 113
      :      :      :      :      :      :      :      :      :      :      :
DB 130 ALLERINHFRTGSIIELLATTAVNCFLPFYQDMPES--ISAQIXMGLINRYKHFSSIPRGF 188
      :      :      :      :      :      :      :      :      :      :      :
QY 114 WLPELAYDPIIPAILKDNQYEVLFADGGAMLFSAHL 149
      :      :      :      :      :      :      :      :      :      :      :
DB 189 YLPELGYALERTIKSYGFSYILETHSFLEGRV 224
      :      :      :      :      :      :      :      :      :      :      :

RESULT 15
Q97ZD2
ID Q97ZD2 PRELIMINARY; PRT: 447 AA.
AC Q97ZD2;
DT 01-OCT-2001 (TREMBSRel. 18, Created)
DT 01-OCT-2001 (TREMBSRel. 18, Last sequence update)
DT 01-OCT-2001 (TREMBSRel. 18, Last annotation update)
DE ALPHA AMYLASE (EC 3.2.1.1).

```

RESULT	15
Q97ZD2	
ID	Q97ZD2 PRELIMINARY; PRT: 447 AA.
AC	Q97ZD2;
DT	01-OCT-2001 (TReMBElrel. 18, Created)
DT	01-OCT-2001 (TReMBElrel. 18, Last sequence update)
DT	01-OCT-2001 (TReMBElrel. 18, Last annotation update)
DE	ALPHA AMYLASE (EC 3.2.1.1).
GN	SSO0988.
OS	Sulfolobus solfataricus.
OC	Archaea; Crenarchaeota; Sulfolobales; Sulfolobaceae; Sulfolobus.
OX	NCBI_TaxID=2287;
RN	[1]

SEQUENCE FROM N.A.
RC STRAIN-AICC 35092 / DSM 1617 / P2;
RX MEDLINE=21322996; Pubmed=11427726;
RA She Q., Singh R.K., Confalonieri F., Zivanovic Y., Allard G.,
RA Aways M.J., Chan-Weiher C.C.-Y., Clausen I.G., Curtis B.A.,
RA De Moors A., Erasuo G., Fletcher C., Gordon P.M.K.,
RA Heikamp-de Jong I., Jeffries A.C., Kozera C.J., Medina N., Peng X.,
RA Thi-Ngoc H.P., Redder P., Schenk M.E., Theriault C., Tolstrup N.,
RA Charlebois R.L., Doolittle W.F., Duquet M., Gaasterland T.,
RA Garrett R.A., Ragan M.A., Sensen C.W., Van der Oost J.;
RT "The complete genome of the crenarchaeon *Sulfolobus solfataricus* P2. ",
RL Proc. Natl. Acad. Sci. U.S.A. 98:7835-7840(2001).
DR ENBL: AF006718; AAK41260.1; -
DR EMBL: AF006718; AAK41260.1; -
DR Hydrolyase: Glycosidase; Complete proteome.
SQ SEQUENCE 447 AA, 53641 MW, E345F31AC2D20EE CRC64;

Query Match 5.8%; Score 109.5; DB 17; Length 447;
Best Local Similarity 20.8%; Pred. No. 1.2;
Matches 72; Conservative 54; Mismatches 135; Indels 85; Gaps 17;

```

Qy 17 KSEIPKVIKAYIPVITELI-----KEE---IPGCLNTGTYTLKFLP---KDIDLKVG 54
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 41 KEIFERIKKNICYIPATNIIILSSIERAAEECNVNVYFFSISGTFLEQAEKRWGSEIVELFQQ 100
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 65 GIASDLIEIGTGYTHAIPLPL-PLSRVQAQVQOREVKEELFELSPKGFMLPELAYDPI 123
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 101 LAYTHKVEPLAQYIVHSVLSLWEDKSEWKEQVMHKDKTIKSYGGYPTTFENTEL----- 155
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 124 IPAILKDNQVYEYLFADGEAMLSAHLNSAIKPIKPLYPHLIAQAEKREPRYSYLLGLRE 183
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 156 ---ITKKDIVEEYKMGFKMMLSEGTRNMLNGRSPNYVYKLG-HEIRMLFRNYTLS--- 208
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 184 LRRAIKLVFEKGVTLKAVKDIE---APVYVAVNTAVMLGIRGLPLMPKPKVASVIEDK 240
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 209 -----DDIAFRFSPNW-----DOYPL-TASKYADWISRSE 238
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 241 NILLYGTDIEFIGYDIAGRYSVEGLLEVID-----ELNS---ELCLSELKHSGRELV- 292
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 239 G----NVGLIFDYTFEGEHRHROTGILEFKMLPTELNSKGVMEMMPKEVYN---DVID 291
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 293 ----LRTSSWA----PKDS----LRTWREDEGNARLNMLSYNMGRG 326
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 292 EIELAHTTSWADIEKESKWLGNIMOWAYDDAVRAEMPSRELGNR 337
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

Search completed: June 2, 2002, 18:09:36
Job time: 218 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: June 2, 2002, 18:01:42 ; Search time 32.53 Seconds
(without alignments)
1242.879 Million cell updates/sec

Title: US-09-886-400-4
Perfect score: 1877
Sequence: 1 LRALVHGNLQYAEIPKSEI.....RRLDAFRALYNDWRGENCEP 364

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_032802.*
1: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1873	99.8	364	AAW34643	Thermotable alpha
2	145.5	7.8	855	AAW34643	Putative P. abyssi
3	143.5	7.6	649	AAW47504	Pyrococcus furiosus
4	124	6.6	1362	AAW6091	Putative P. abyssi
5	121.5	6.5	653	AAW54870	Super heat resista
6	104.5	5.6	360	AAW82528	S. epidermidis ope
7	104.5	5.6	376	AAW82506	S. epidermidis ope
8	104	5.5	329	AAW4013	Heat resistant mal
9	104	5.5	571	AAW16038	Arabidopsis thalia
10	104	5.5	571	AAW48655	Arabidopsis thalia
11	104	5.5	704	AAW16037	Arabidopsis thalia

12	104	5.5	704	21	AAW48654	Arabidopsis thalia
13	104	5.5	726	21	AAW16036	Arabidopsis thalia
14	104	5.5	726	21	AAW48653	Arabidopsis thalia
15	103.5	5.5	936	22	ABG18270	Novel human diagno
16	97	5.2	929	22	ABW70857	Drosophila melanog
17	96	5.1	1073	22	AAW36758	Staphylococcus aur
18	96	5.1	1147	22	AAU33972	Staphylococcus aur
19	93.5	5.0	317	22	AAW96470	Putative P. abyssi
20	92.5	4.9	480	21	AAW16318	Pinus radiata UGP
21	92.5	4.9	1668	22	AAU34158	Staphylococcus aur
22	92.5	4.9	2397	22	AAU36672	Staphylococcus aur
23	92	4.9	216	21	AAW29078	Arabidopsis thalia
24	92	4.9	218	21	AAW29077	Arabidopsis thalia
25	92	4.9	235	21	AAW29076	Arabidopsis thalia
26	92	4.9	1417	20	AAW31947	Plasmodium falcipa
27	91.5	4.9	473	21	AAW07943	A uridine diphosph
28	91.5	4.9	473	21	AAW34044	Zea mays protein f
29	91	4.8	364	21	AAW14039	Arabidopsis thalia
30	91	4.8	375	22	AAW82373	S. epidermidis ope
31	91	4.8	389	21	AAW14038	Arabidopsis thalia
32	91	4.8	852	22	AAW15556	Corynebacterium gl
33	91	4.8	852	22	AAW33245	C glutamicum prote
34	91	4.8	4829	22	AAW97833	Human apoptosis-in
35	90.5	4.8	377	21	AAW34045	Zea mays protein f
36	90	4.8	844	21	AAW36516	Candida albicans C
37	90	4.8	1039	22	ABG15145	Novel human diagno
38	90	4.8	1248	22	ABW61465	Drosophila melanog
39	89.5	4.8	1057	19	AAW72067	HSV-2 strain SB5 C
40	89.5	4.8	1114	19	AAW72205	HSV-2 strain SB5 C
41	89.5	4.8	1114	19	AAW72096	HSV-2 strain SB5 C
42	89.5	4.8	3854	22	ABW67104	Drosophila melanog
43	89.5	4.8	5385	22	ABW66487	Drosophila melanog
44	89.5	4.8	5496	22	ABW67161	Drosophila melanog
45	89.5	4.8	8805	22	ABW67112	Drosophila melanog

ALIGNMENTS

RESULT 1
AAW34643
ID AAW34643 standard; Protein; 364 AA.
XX
AC AAW34643;
DT 27-MAR-1998 (first entry)
XX
DE Thermotable alpha-galactosidase AEDII12RA-alpha-gal-189C.
XX
KW Alpha-galactosidase; alpha-glycosidase; thermostable enzyme;
KW food processing; alpha glycoside hydrolysis; raffinose;
KW stachyose; verbascose; bean; flatulence; AEDII12RA-alpha-gal-189C.
XX
OS Thermococcus alcaliphilus strain AEDII12RA.
XX
FH Key Location/Qualifiers
FT Misc-difference 329 /note= "encoded by CIT"
FT
XX
PN W09732974-A1.
XX
PD 12-SEP-1997.
XX
PF 05-FEB-1997; 97WO-US01452.
XX
PR 08-MAR-1996; 96US-0613220.
XX
PA (RECO-) RECOMBINANT BIOCATALYSIS INC.
XX
PI Murphy D, Reid J, Rudolph MJ;
XX
DR WPI: 1997-470541/43.
DR N-PSDB; AAT93753.

XX Nucleic acid encoding alpha-galactosidase from Thermococcus
PT alcaliphilus - used in food processing to hydrolyse
PT alpha-galactosides, e.g. raffinose
XX
PS Claim 1; Fig 1; 32pp; English.
XX
CC This protein comprises AEDII12RA-alpha-gal-18GC, a claimed
CC thermostable alpha-galactosidase of Thermococcus alcaliphilus
CC AEDII12RA, a bacterium that shows optimum growth at 85 deg C and
CC pH 9.5. Also claimed are: (1) an isolated polynucleotide (see
CC AAR93753) encoding the alpha-galactosidase; (2) a vector containing
CC the polynucleotide or homologous or complementary sequences; (2)
CC host cells containing the vector; (3) a process for producing the
CC alpha-galactosidase in transformed or transfected host cells; an
CC enzyme showing at least 70% identity to alpha-galactosidase and
CC comprising at least 30 amino acid residues of its sequence; and (4)
CC a method for hydrolysing alpha-galactoside bonds using the enzyme.
CC The enzyme can be used to hydrolyse raffinose to sucrose and glucose
CC in sugar beet processing (raffinose inhibits crystallisation of
CC sucrose), and as a digestive aid to hydrolyse raffinose, stachyose
CC and verbascone in beans and other gassy foods.
XX
SQ Sequence 364 AA;

Query Match 99.8%; Score 1873; DB 18; Length 364;
Best Local Similarity 99.7%; Pred. No. 1.8e-182;
Matches 363; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 LRAIVFGNLOQYAEIPKSEIPKVIKRAYIPVETLKEEIPFGLNITGYTLKFLPKDIID 60
DB 1 LRAIVFGNLOQYAEIPKSEIPKVIKRAYIPVETLKEEIPFGLNITGYTLKFLPKDIID 60
QY 61 LVKGIGASDLTEIIGTSTYTHAILPLSLRVEAQVORDREVKEELFELSPKGFWLPDELAY 120
DB 61 LVKGIGASDLTEIIGTSTYTHAILPLSLRVEAQVORDREVKEELFELSPKGFWLPDELAY 120
QY 121 DPTIPAILKONGYLYLFDAGEMLFSAHLNSAIKPKIPLYPHLIKAQREKRYISYLLG 180
DB 121 DPTIPAILKONGYLYLFDAGEMLFSAHLNSAIKPKIPLYPHLIKAQREKRYISYLLG 180
QY 181 LRELRKAITLVFEKQVTLKAVKDIKAVPVVAVNTAVMLGIGRLPLMNPKKVASWIEDK 240
DB 181 LRELRKAITLVFEKQVTLKAVKDIKAVPVVAVNTAVMLGIGRLPLMNPKKVASWIEDK 240
QY 241 NILLYGTDIEFIGYRDIAGRMSVEGLLEVIDENSELCLPSELKHSGRELYLRTSSWAP 300
DB 241 NILLYGTDIEFIGYRDIAGRMSVEGLLEVIDENSELCLPSELKHSGRELYLRTSSWAP 300
QY 301 DKSRLRWREDEGNARLNMLSYNNMGEGLALLAENSADARGWEPLPERRIDAFRAIYNDRGE 360
DB 301 DKSRLRWREDEGNARLNMLSYNNMGEGLALLAENSADARGWEPLPERRIDAFRAIYNDRGE 360
QY 361 NGEPL 364
DB 361 ngepl 364

RESULT 2
AAB96088
ID AAB96088 standard; Protein; 655 AA.
XX
AC AAB96088;
XX
DT 29-OCT-2001 (first entry)
XX
DE Putative P. abyssi alpha-amylase.
XX
KW Hyperthermophilic archaeon; hyperthermophilic protein.
XX
OS Pyrococcus abyssi.
XX

PN FR2792651-A1.
XX
PD 27-OCT-2000.
XX
PF 21-APR-1999; 99ER-0005034.
XX
PR 21-APR-1999; 99ER-0005034.
XX
PA (CNRS) CNRS CENT NAT RECH SCI.
PA (IFRE-) IFREMER INST FR RECH EXPL MER.
XX
PI Forterre P, Thierry JC, Prieur D, Dietrich J, Lecompte O;
PI Querellou J, Weissenbach J, Saurin W, Heilig R;
XX
DR WPI; 2001-126236/14.
XX
PT New nucleotide sequences isolated from Pyrococcus abyssi encode
PT proteins useful in industry -
XX
PS Claim 7; Pages 711-712; 1657pp; French.
XX
CC The present invention relates to the genomic sequence of Pyrococcus
CC abyssi (see AAR86431 and AAR41223-7) and P. abyssi proteins. P. abyssi is
CC a hyperthermophilic archaeon, which is isolated from deep-sea
CC hydrothermal vents. The present sequence is one such P. abyssi protein.
CC The proteins of the present invention have various potential industrial
CC uses, since the proteins are stable at very high temperatures, some up to
CC 110 degrees centigrade.
CC Note: This patent is in the same patent family as WO2000065062, which
CC contains additional sequences as shown in AAB99132-AAB99143,
CC AAR75903-AAR75920 and AAG66436.
XX
SQ Sequence 655 AA;

Query Match 7.8%; Score 145.5; DB 22; Length 655;
Best Local Similarity 22.4%; Pred. No. 1.1e-05;
Matches 90; Conservative 55; Mismatches 127; Indels 129; Gaps 22;
QY 25 EKAVIPVETLKEEIP---FGLNITGYTLKFLPK---DIIDLKGGIASDLTEIIGTSTY 78
DB 28 ekayrpfleil--eeypnmkvaihsgilvewleenkpyidllkslvrkqgveivvvgf 85
QY 79 THAILPLSLRVEAQVORDREVKEELFELSPKGFWLPDELAYDPTIPAILKONGYLYLFA 138
DB 86 yepvlaaipkedrleciyllkewakki-gydakglwtervqgpeivktlreagleyvvv 144
QY 139 DGEAMLFSAHLSA-IKPIKPLYPHLIKAQRE-----KRPY-----ISILL 179
DB 145 d-----dyhmsaglskdqlfwpyytedggevityfp.deklrylipfrpdkvisylh 198
QY 180 GLRELKAIKLVFSGKVTLKAVKDIEAVPVVAVNTAVMLGIGRLPLMNPKKVASWIEDK 239
DB 199 slasedskvavfh-----ddgekfgiw-----pm-----tyewvyek 231
QY 240 D-----NILLYGTDIE-----FIGYRDIAGY-----RMSVEG 266
DB 232 gwlreffdrvssdeainimlyseylqkfkpgkglvlypiasyfemsewslpaggaklfe- 290
QY 267 LLEVIDENSELCLPSELKHSGRELYLRTSSWAPDKSLRWREDEGN---ARLNMLSYNN 323
DB 291 fveklkeln-----mferyrvfvggiw---kxf-fykypeanymhkmmlms---- 334
QY 324 RGEALLAENSADARGWEPLPERRIDAFRAIYNDRGE 362
DB 335 ---rlirdpsarrf-----vraqcndaywhgvfg 362

RESULT 3
AAR47504
ID AAR47504 standard; Protein; 649 AA.
XX
AC AAR47504;

```

XX 07-JUL-1994 (first entry)
XX Pyrococcus furiosus alpha amylase.
XX Pyrococcus furiosus; alpha amylase; liquefaction; polymers;
XX glucopolymers; thermostable.
XX Pyrococcus furiosus.
XX EP577257-A.
XX 05-JAN-1994.
XX 17-MAY-1993; 93EP-0303801.
XX 09-JUN-1992; 92US-0893928.
XX (UYJO ) UNIV JOHNS HOPKINS.
XX Anfinsen CB, Laderman K;
XX WPI; 1994-009532/02.
XX Purified Pyrococcus furiosus alpha-amylase - used for the
XX industrial liquefaction of gluco-polymers at high temps.
XX Claim 2; Figure 9; 4lpp; English.
XX The purified Pyrococcus furiosus alpha amylase can act on substrates
XX with a low degree of polymerisation. e.g. glucose polymers as short
XX as maltotriose. The enzyme can be used for efficient industrial
XX liquefaction of glucopolymers at high temperatures.
XX Sequence 649 AA;

Query Match 7.6%; Score 143.5; DB 15; Length 649;
Best Local Similarity 22.8%; Pred. No. 1.8e-05;
Matches 91; Conservative 59; Mismatches 118; Indels 131; Gaps 24;

QY 25 EKAYIPVETLKEEIP---FGLNITGYTLKPLPKDI---IDLKGGIASDLIEIGTS 77
DB 29 ekcywplfletl--eeypmkxvahstgpliewl-qdirpeyldilrslvkrqgvelvvg 85

QY 78 YTHAILPLPLSLRVEAQVORDREVKE--ELFELSPKGFWLPPELAYDPIIPAILKDNQY 135
DB 86 fyeplvasip--kedrieglrmkewaksigfdargvwltervwqpelvktlkesgidy 142

QY 136 LFADGEMLFSAHLNS-----AIXPIKPLYPHLIKAOPEKRFYISVLL 179
DB 143 viyvd-gyhfmsaelskeelywpyytedggevavipid-----ek-----Iryli 186

QY 180 GLRELKAKL---VPEGVTKAV--KDIEAVPVVWAVNTAVMLGIGRLPLMNPKKVAS 234
DB 187 pfrpvckvieylhslidgdeskvavfhdggekfgiwpgtiewvy-----ekg 233

QY 235 WI-----EDKNILLYGTDIE-----FCGYRDIAGYRMSVEGLEVDLSELCL 280
DB 234 wlreffdrissdeklnlmlyteylekyxprgivyipiasy-----fem-----sewsl 281

QY 281 P-----SELKHS-----RELYLRTSSWAPDKSLRIWREDEGN---ARLNMJSYN 322
DB 282 pakqarfvefnelkvkifekyrvfvggiw---knf-fykypesnymhkrmlmvskl 337

QY 323 MRGELALLAENSARGWEPLPERRLDFAFIYND--WRG 359
DB 338 vrnn-----pearkyllraqcndwvhg 360

RESULT 4
AAB96091
ID AAB96091 standard; Protein; 1362 AA.

```

```

XX AAB96091;
XX 29-OCT-2001 (first entry)
XX Putative P. abyssi amylopullulanase.
XX Hyperthermophilic archaeon; hyperthermophilic protein.
XX Pyrococcus abyssi.
XX FR2792651-A1.
XX 27-OCT-2000.
XX 21-APR-1999; 99FR-0005034.
XX 21-APR-1999; 99FR-0005034.
XX (CNES ) CNRS CENT NAT RECH SCI.
XX (IFRE-) IFREMER INST FR RECH EXPL MER.
XX Forterre P, Thierry JC, Prieur D, Dietrich J, Lecompte O;
XX Querellou J, Weissenbach J, Saurin W, Heilig R;
XX WPI; 2001-126236/14.
XX New nucleotide sequences isolated from Pyrococcus abyssi encode
XX proteins useful in industry -
XX Claim 7; Pages 715-719; 1657pp; French.
XX The present invention relates to the genomic sequence of Pyrococcus
XX abyssi (see AAF8431 and AAH41223-7) and P. abyssi proteins. P. abyssi is
XX a hyperthermophilic archaeon, which is isolated from deep-sea
XX hydrothermal vents. The present sequence is one such P. abyssi protein.
XX The proteins of the present invention have various potential industrial
XX uses, since the proteins are stable at very high temperatures, some up to
XX 110 degrees centigrade.
XX Note: This patent is in the same patent family as WO200065062, which
XX contains additional sequences as shown in AAB99132-AAB99143,
XX AAH75903-AAH75920 and AAG66436.
XX Sequence 1362 AA;

Query Match 6.6%; Score 124; DB 22; Length 1362;
Best Local Similarity 21.5%; Pred. No. 0.0056;
Matches 82; Conservative 57; Mismatches 142; Indels 100; Gaps 16;

QY 32 IETLIKEEIPFGLNITGYTLKPLPKDIIDLKGGIASDLIEIGTSYTHAILPLP--- 87
DB 236 vetvlkqgm-wlnht-----feeheklnlllg---ngnvetvvtvpythpigliindfgw 286

QY 88 LSRVEAQVORDREVKEELF---ELSPKGFWLPPELAYDPIIPAILKDNQYELFADG---E 141
DB 287 yedfdagvkkanelykeylgagkvtpkggwaesaalndkt-eilaengkwkwmtdqivle 346

QY 142 AMLFSAHLNSAKP-----IKPLYPHLIKA 166
DB 347 klgvptkiesytkpwwagfgdkkiylfprnhlsdrvfyagmnqydavknfveeliki 406

QY 167 QRKRFPRYSYLIGLRELKAKLKFEGKVTLKAV-KDIEAVPVVWAVNTAV-----ML 219
DB 407 qknydgsviyvitidgenpwehyfdgklfleelyrgleekqkglirtvtpsey-emf 466

QY 220 G--IGRLPLMNPKKVASWIEDKDNILLYGTDFIEFGYRDIAGYRMSVEGLEVDLSE 277
DB 467 gdkankltpkmmkridftteddnvallkaktl-----yelydmvgvte----- 509

QY 278 LCPLSELKHSRELYLRTSSWAPDKSLRIWREDEGNARLNMJSYNMRGELALLAENS 336
DB 510 -----emqwpesswi-dgtlstw-----igepgeniawwylyarkalfenkdnv 553

```

QY 337 RGWEPLPERRLDARAIYNDK 357
Db 554 kwnkayey---lfragsdw 571
RESULT 5
AAW54870
ID AAW54870 standard; Protein; 653 AA.
XX AAW54870;
AC AAW54870;
DT 01-SEP-1998 (first entry)
DE Super heat resistant 4-alpha-glucanotransferase.
KW Super heat-resistant 4-alpha-glucanotransferase; heat-treatment;
KW alpha-1,4-glucan; alpha-1,4-glucoside bond.
OS Pyrococcus sp.
PN JP10150986-A.
PD 09-JUN-1998.
PF 21-NOV-1996; 96JP-0311117.
XX 21-NOV-1996; 96JP-0311117.
PR (BEAB-) BE ABLE KK.
PA (NAGA-) NAGASE SEIRAGAKU KOGYO KK.
XX WPI: 1998-379989/33.
DR N-PSDB; RAV27026.
XX New 4-alpha-glucanotransferase which has been heat-treated
PT high temperatures to transfer at least one glucose unit
XX
PS Claim 4; Page 10-13; 18pp; Japanese.
CC The super heat-resistant 4-alpha-glucanotransferase has an optimum pH of
CC 6.0-8.0. with an optimum temperature at pH 7.5 of 100 deg. C. It has at
CC least 90% activity after heat-treatment at 100 deg. C for 30 minutes at
CC pH 7.5. It can transfer at least one glucose unit in alpha-1,4-glucan
CC to alpha-1,4-glucan by an alpha-1,4-glucoside bond.
XX
SQ Sequence 653 AA;

Query Match 6.5%; Score 121.5; DB 19; Length 653;
Best Local Similarity 21.2%; Pred. No. 0.0032;
Matches 82; Conservative 64; Mismatches 141; Indels 99; Gaps 21;
QY 25 EKAYIPVETTLKKEIP---PGLNITGYTLKFL---PKDIIDLKVGSIASDLIEIGTSY 78
Db 28 ersyrfmetl--eeypmkvavhygsppliewrdnkpehldllslvkrqgleivvagf 85
QY 79 THAILPLPLSRVQVQORDREVKEELFELSPKGFVLPYLAOYPIPAIKKONGYELFA 138
Db 86 yepviasipkedrivqieklkefarnl-gyeardgwiterwvqgelvkslraagidyviv 144
QY 139 DGEAMLFSAHLNSA-IKPIKPLYPHLIKAQRE-----KRFRYSYLLGLRELKRAIK 189
Db 145 d-----dyhfmagsqiskdelvfpyytedggevityvfpideklr---ylipfrpvaktie 195
QY 190 LVF---EGKVTLKAV--KDIEAVPWPVAVNTAVMLGIGRLPLMNPKNKVASWI----- 236
Db 196 yhsiddgdeskvavfhdgdegkfgwptgyewy-----ekgwlrfeffdrvs 242
QY 237 -EDKNILLYGTDIE----FIGYRDYAGY-RMSVEGL-----LEVIDELNSELCLP 281
Db 243 sderinlmlyseylrfrprgrglvipyasiyfemsewslparqaklfvfeveelk----- 296

QY 282 SELKHSORELYLRTSSWAPDKSLRIWREDEGN---ARLNMLCYNMRGELALLAENSARG 338
Db 297 kenkfdryrvfrvggiw---knf-fkypesnymkhkrmmlmvskavrnn----- 340
QY 339 WEPLPERRLDARAIYND--WRGENG 362
Db 341 ----pearefilraqcndaywhgvfg 362
RESULT 6
AAG82528
ID AAG82528 standard; Protein; 360 AA.
XX AAG82528;
AC AAG82528;
DT 03-SEP-2001 (first entry)
DE S. epidermidis open reading frame protein sequence SEQ ID NO:2150.
KW Staphylococcus epidermidis SRI strain; infection; diagnosis;
KW vaccination; endocarditis.
OS Staphylococcus epidermidis.
PN WO200134809-A2.
PD 17-MAY-2001.
XX 09-NOV-2000; 2000WO-US30782.
PF 09-NOV-1999; 99US-0164258.
PR (GLAX) GLAXO GROUP LTD.
XX Kimmerly WJ;
XX WPI: 2001-316495/33.
DR N-PSDB; AAH53378.
XX Nucleic acids encoding polypeptides from Staphylococcus epidermidis,
XX useful for vaccinating against infections, e.g. endocarditis -
XX Claim 18; Page 578-579; 2188pp; English.
AAH52304 to AAH53970 represent nucleic acids (I) encoding polypeptides
(II), given in AAG81454 to AAG83120, from Staphylococcus epidermidis.
(I) and (II) can have antibacterial activity and therefore can be used
in vaccination. The nucleic acids (I) may be used to produce the
S. epidermidis polypeptides (II) via the production of vectors
containing them which are used to produce hosts cells which express the
polypeptides. The polypeptides (II) (and/or nucleic acids) may then be
used to vaccinate subjects and to raise antibodies against the bacteria.
The polypeptides may also be used to assay for other inhibitors of their
activity and therefore identify compounds that may be used for the
treatment of S. epidermidis infections, e.g. endocarditis. AAH53971 to
AAH55090 represent specifically claimed S. epidermidis genomic DNA
polynucleotide sequences from the present invention. AAH55091 to
AAH55098 represent oligonucleotide sequences and primers which are used
in the exemplification of the present invention.
N.B. The present invention specifically claims all the polynucleotide
sequences given in the sequence listing of the present specification,
however the sequence listing only goes up to SEQ ID NO:4454 so even
though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,
no sequences are present for SEQ ID NO:4455 to 4464.
XX Sequence 360 AA;
SQ

Query Match 5.6%; Score 104.5; DB 22; Length 360;
Best Local Similarity 19.4%; Pred. No. 0.069;
Matches 58; Conservative 51; Mismatches 117; Indels 73; Gaps 12;
QY 4 LVFHGNLOYAIPKSEIPKVIKAYIPIVITLKEEIPFGLNITGYTLKPLPKDIDLK 63

```

CC though sequences are given in the disclosure for SEQ ID NO:4465 to 4472,
XX no sequences are present for SEQ ID NO:4455 to 4464.
SQ Sequence 376 AA;

Query Match          5,68; Score 104.5; DB 22; Length 376;
Best Local Similarity 19.48; Pred. No. 0.074;
Matches 58; Conservative 51; Mismatches 117; Indels 73; Gaps 12

QY      4 LVFHGNLOAYEIPKSEIKVIEKAYIPVIETLIIKEEIPFGLNITGYTLKFLPKDIIIDLVK 63
       | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db     47 lifkg-----vkrivedgygsiirklikunl--nlialhtnldvnpgkgnmlia 93

QY      64 GGIASDLTEIGT--SYTHALPLPLISRVFAQVORDREVKEELFELSPKGFWLPCLAYD 121
       | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db     94 dqlglenlsmitntssyykvqtfrpknyie-----dfkdslnel----- 133

QY     122 PIPAILKDNGVEYLFDAGEAMLFSAHLNSAIKPIKPLYPHLIKAQRKFRFYISYLLGL 181
       | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db    134 ----glakegnyeycfeseg-----kgqfcpvgdaspyigklde-----leyv--- 173

QY     182 RELRAIKLVF-----EGKVILKAVKDIE--AVPV--WVAIVTAVMLGIRGLPMLNPKKV 232
       ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db    174 ----deiklefmiakndneleitraildnhpetyetpvfdikmkeseyvigilgblqmt 229

QY     233 ASWIEDKNILLYGTDIFIGVRDIAGYRMSVEGLEEV-----IDELNSELCIPLSELKH 286
       | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db    230 ldefseyakkqlinpsrvtytghdspikkvaillggsgifgeykasqlgadqvftgdikh 288

RESULT 8
AAR94013
ID AAR94013 standard; Protein; 329 AA.
XX
AC AAR94013;
XX
DT 21-AUG-1996 (first entry)
XX
DE Heat resistant maleate dehydrogenase.
XX
KW Heat resistant maleate dehydrogenase; h-RMAD; NADH; L-aspartic acid;
KW alpha-ketoglutaric acid; glutamine oxalo-transaminase activity; GOT.
XX
OS Bacillus stearothermophilus ATCC 12016.
XX
Key Location/Qualifiers
FT Misc-difference 185
FT /note= "Given in the specification as Var"
XX
PN JP08047389-A.
XX
PD 20-FEB-1996.
XX
XX 01-JUL-1994; 94JP-0151045.
XX
PR 03-JUN-1994; 94JP-0121629.
PR 02-JUL-1993; 93JP-0164701.
XX
XX {TOYM ) TOYOBO KK.
XX
XX WPI: 1996-166248/17.
DR N-PDSB; AAT17715.
XX
PT Protein having heat resistant maleate dehydrogenase activity - and
PT reagent contg. protein, NADH and L-aspartic and alpha-keto:glutaric
PT acid for determination of glutamine oxalo-transaminase activity
XX
PS Claim 8; Page 13-15; 17pp; Japanese.
XX
XX This sequence represents a protein having heat resistant maleate
XX dehydrogenase (h-RMAD) activity. The protein has a residual activity
CC after storage at 40 deg.C for 10 days of at least 60%, pret 70%

```

[illegible]

PR	99US-0136782
PR	99US-0137222
PR	99US-0137528
PR	99US-0137502
PR	99US-0137724
PR	99US-0138094
PR	99US-0138540
PR	99US-0138684
PR	99US-0139119
PR	99US-0139452
PR	99US-0139453
PR	99US-0139492
PR	99US-0139454
PR	99US-0139455
PR	99US-0139456
PR	99US-0139457
PR	99US-0139458
PR	99US-0139459
PR	99US-0139460
PR	99US-0139461
PR	99US-0139462
PR	99US-0139463
PR	99US-0139464
PR	99US-0139465
PR	99US-0139466
PR	99US-0139467
PR	99US-0139468
PR	99US-0139469
PR	99US-0139470
PR	99US-0139471
PR	99US-0139472
PR	99US-0139473
PR	99US-0139474
PR	99US-0139475
PR	99US-0139476
PR	99US-0139477
PR	99US-0139478
PR	99US-0139479
PR	99US-0139480
PR	99US-0139481
PR	99US-0139482
PR	99US-0139483
PR	99US-0139484
PR	99US-0139485
PR	99US-0139486
PR	99US-0139487
PR	99US-0139488
PR	99US-0139489
PR	99US-0139490
PR	99US-0139491
PR	99US-0139492
PR	99US-0139493
PR	99US-0139494
PR	99US-0139495
PR	99US-0139496
PR	99US-0139497
PR	99US-0139498
PR	99US-0139499
PR	99US-0139500
PR	99US-0139501
PR	99US-0139502
PR	99US-0139503
PR	99US-0139504
PR	99US-0139505
PR	99US-0139506
PR	99US-0139507
PR	99US-0139508
PR	99US-0139509
PR	99US-0139510
PR	99US-0139511
PR	99US-0139512
PR	99US-0139513
PR	99US-0139514
PR	99US-0139515
PR	99US-0139516
PR	99US-0139517
PR	99US-0139518
PR	99US-0139519
PR	99US-0139520
PR	99US-0139521
PR	99US-0139522
PR	99US-0139523
PR	99US-0139524
PR	99US-0139525
PR	99US-0139526
PR	99US-0139527
PR	99US-0139528
PR	99US-0139529
PR	99US-0139530
PR	99US-0139531
PR	99US-0139532
PR	99US-0139533
PR	99US-0139534
PR	99US-0139535
PR	99US-0139536
PR	99US-0139537
PR	99US-0139538
PR	99US-0139539
PR	99US-0139540
PR	99US-0139541
PR	99US-0139542
PR	99US-0139543
PR	99US-0139544
PR	99US-0139545
PR	99US-0139546
PR	99US-0139547
PR	99US-0139548
PR	99US-0139549
PR	99US-0139550
PR	99US-0139551
PR	99US-0139552
PR	99US-0139553
PR	99US-0139554
PR	99US-0139555
PR	99US-0139556
PR	99US-0139557
PR	99US-0139558
PR	99US-0139559
PR	99US-0139560
PR	99US-0139561
PR	99US-0139562
PR	99US-0139563
PR	99US-0139564
PR	99US-0139565
PR	99US-0139566
PR	99US-0139567
PR	99US-0139568
PR	99US-0139569
PR	99US-0139570
PR	99US-0139571
PR	99US-0139572
PR	99US-0139573
PR	99US-0139574
PR	99US-0139575
PR	99US-0139576
PR	99US-0139577
PR	99US-0139578
PR	99US-0139579
PR	99US-0139580
PR	99US-0139581
PR	99US-0139582
PR	99US-0139583
PR	99US-0139584
PR	99US-0139585
PR	99US-0139586
PR	99US-0139587
PR	99US-0139588
PR	99US-0139589
PR	99US-0139590
PR	99US-0139591
PR	99US-0139592
PR	99US

PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.

PR	18-JUN-1999	90S-0139763
PR	21-JUN-1999	90S-0139817
PR	22-JUN-1999	90S-0139859
PR	23-JUN-1999	90S-0140333
PR	23-JUN-1999	90S-0140354
PR	24-JUN-1999	90S-0140695
PR	28-JUN-1999	90S-0140823

99US-0142154.
99US-0142055.
99US-0142390.

PR 12-JUL-1999; 99US-0142377
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.

PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.

PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.

PR	22-JUL-1999;	99US-0145085.
PR	22-JUL-1999;	99US-0145087.
PR	22-JUL-1999;	99US-0145089.

PR	23-JUL-1999;	99US-0145224.
PR	26-JUL-1999;	99US-0145276.

```
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 28-JUL-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147439.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154770.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157755.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 21-OCT-1999; 99US-0160816.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.

PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161921.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 5.5%; Score 104; DB 21; Length 571;
Best Local Similarity 22.5%; Pred. No. 0.16;
Matches 82; Conservative 49; Mismatches 112; Indels 122; Gaps 20;

QY 29 IPVETLIKEIPGLNITGYTLKFLPKDIDLVKGGIADL-----IEIIGSYTHAIL 83
Db 101 ipile-mvpeelhf---tvpvalsfimgdpkmatlgidbqlptgvkiekrlrlktml 156
QY 84 PLLPLSRVEAQVQRD-----RVKDELFEL-SPKGFWLPPELAYDP 122
Db 157 pl--lseiigiipretllwklklrrscayansrihavqaeivlasgkdmmlpsqeeak 214
QY 123 IIPAIL-----KNGYEYLFADGEAMLFSAHLSNAIKPIPLPHILIKAOEKRFY 174
Db 215 rihgilkncsvrcfkdnghltlledsisll-----tvikgtg-----kyrrswrydl 261
QY 175 IS-----YLLG--LRELKKAIKLVF-----EGKVTLRKAVKDIEAVP---VWVAV 213
Db 262 vsdfilppskgealaydevlgflrnavgsvffstmedgki-----vkglagvpdkgpvlly 317
QY 214 NTAVMLGIGRLPLMNPKKVASWIEDKNILL-----YGTDFIEFIGY 254
Db 318 gyhlmgliekgmse-----afikek-nilfrgmahpvlvsdndpaxafdygdwkvfvg- 370
QY 255 RDIAGYRMSVEGLLEVIDELNSCLPS---ELKHSGRELYLRTSSWAPDKSLRWREDE 311
Db 371 ----ayvtatnlfkildskshvllfpogarealnhrgey-----kllwpeqq 415
QY 312 GNARL 316
Db 416. efvrn 420

RESULT 10
AAG48655
ID AAG48655 standard; Protein: 571 AA.
XX
AC AAG48655;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 61465.
XX
KW protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
PF 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
```

PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 04-MAY-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134321.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 14-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139817.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140951.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 28-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148241.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 22-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.

Db 549 efvm 553
|:
RESULT 12
AAG48654
ID AAG48654 standard; Protein; 704 AA.
XX
AC AAG48654;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 61464.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
hybridisation assay; genetic mapping; gene expression control; promoter;
termination sequence.
KW
XX
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
PR 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 28-APR-1999; 99US-0130891.
PR 30-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 11-MAY-1999; 99US-0132863.
PR 14-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 02-AUG-1999; 99US-0147038.
PR 03-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.

```
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149729.
PR 23-AUG-1999; 99US-0149920.
PR 23-AUG-1999; 99US-0149930.
PR 23-AUG-1999; 99US-0150566.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 28-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match      5.5%; Score 104; DB 21; Length 704;
Best Local Similarity 22.5%; Pred. No. 0.22;
Matches 82; Conservative 49; Mismatches 112; Indels 122; Gaps 20;

QY 29 IPVETLKEIPGLNITGYTKFLPKDIDLVKGGIASDL-----IEIGTSYTHAIL 83
Db 234 lpile-mvpeelhf---tvpysfmgdbpkmatidgiclpqgvkieklrqltkml 289
QY 84 PLLPLSRVEAOVQD-----REVKEELFEL-SPKGFVLPPELAYDP 122
Db 250 pl--lseiggiptrellwlklrrscayansrihavqaeivlasgkdmmlpsqeak 347
```

```
QY 123 IIPAIL-----KDNQYEYLFADGEAMLFSAHLSAIAKPKLYPLHLKAAQREKFRY 174
Db 348 rilgllkncsvrcfkdnghltlledsisil-----tvikgtg-----kyrrswydl 394
QY 175 IS-----YLLG--LRELKAIKLVF-----EGKVTLKAVKDIEAVP---VWVAV 213
Db 395 vsdflppskegiayaldevlgfmrnavsvffstmedeki---vkglagvpdkgpvlly 450
QY 214 NTAVMLGIGRLPLMNPKNVSAWIEDKNILL-----YGTDFEFIGY 254
Db 451 gyhmlmglelqpmse-----afikek-nilfrgmahpvlvsndnpakafdygdwkwfvg- 503
QY 255 RDTAGYRMSVEGLEVIDELNSELCLPS--ELKHSGRLEYLRTSSWAPDKSLRIWREDE 311
Db 504 ----aypvtatnlfkildsksnvlfpqgarealhrgeqy-----kllwpeqq 548
QY 312 GNARL 316
Db 549 efvrn 553

RESULT 13
AAG16036
ID AAG16036 standard; Protein; 726 AA.
AC AAG16036;
XX
XX
DT 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 16522.
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
OS Arabidopsis thaliana.
XX
PN EP1033405-A2.
XX
PD 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
XX 25-FEB-1999; 99US-0121825.
PR 03-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0126264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 07-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0132863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
```

PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135625.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136399.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138847.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0141554.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147036.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 09-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149929.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 30-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0159293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 28-OCT-1999; 99US-0161360.
PR 28-OCT-1999; 99US-0161361.

```
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match      5.5%; Score 104; DB 21; Length 726;
Best Local Similarity 22.5%; Pred. No. 0.23;
Matches 88; Conservative 49; Mismatches 112; Indels 122; Gaps 20;

QY 29 IPVTETLIKKEIPLGELITKFLPKDIDLVKGGIASDL-----IILIGTSYTHAIL 83
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 256 lpile-mvpeelhf--tvpyaisfmgdpkmatigndqptgkiskirgrtktml 311
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 84 PLLPLSVEAQVORD-----REVKEELFEL-SPRGFWLPELAYDP 122
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 312 pl-lselggiipretllwkllrsgcayansrihavqaevlvlasgkdmmlpsqeak 369
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 123 IIPAIL-----KONGEYELFADGEAMLFSAHLNSAKPKPLYPHLIKAQREKRFY 174
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 370 rlhglkncsvrcfdnghtllledsisll-----tvikgtg-----kyriswrydl 416
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 175 IS-----YLLG--LRELKAIKLVF-----EGKVTLKAVKDIEAVP---VWVAV 213
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 417 vsdfllpskgealayadevlgfirnavgsvffstmedgki---vkglagvpdkgpvlly 472
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 214 NTAYMLGIGRLPLMNPXKVASWIEDKDNILL-----YGTDIEFIQY 254
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 473 gyhnlmgieigpmse-----afikek-nilfrgmahpvlvsndndpakafdygdwkvfg- 525
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 255 RDIAGYRMSVEGLLEVIDENSELCLPS---ELKHSGRELYLTSSWAPDKSLRIWREDE 311
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 526 ----aypytatnlfkldskshvllfpggarealhrgesy-----kllwpeq 570
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 312 GNARL 316
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 571 efvrn 575

RESULT 14
AAG48653
ID AAG48653 standard; Protein; 726 AA.
XX
AC AAG48653;
XX
DT 18-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 61463.
XX
KW Protein identification; signal transduction pathway; metabolic pathway;
KW hybridisation assay; genetic mapping; gene expression control; promoter;
KW termination sequence.
XX
OS Arabidopsis thaliana.
PN EP1033405-A2.
PD
PF 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
XX 25-FEB-1999; 99US-0121825.
XX 05-MAR-1999; 99US-0123180.
XX 09-MAR-1999; 99US-0123548.
XX 23-MAR-1999; 99US-0125788.
XX 25-MAR-1999; 99US-0126264.
XX 29-MAR-1999; 99US-0126785.
XX 01-APR-1999; 99US-0127462.
XX 06-APR-1999; 99US-0128234.
XX 08-APR-1999; 99US-0128714.
XX 16-APR-1999; 99US-0129845.
XX 19-APR-1999; 99US-0130077.
XX 21-APR-1999; 99US-0130449.
XX 23-APR-1999; 99US-0130510.
99US-0130891.
99US-0131449.
99US-0132048.
99US-0132407.
99US-0132484.
99US-0132485.
99US-0132486.
99US-0132487.
99US-0132863.
99US-0134256.
99US-0134218.
99US-0134219.
99US-0134370.
99US-0134768.
99US-0134941.
99US-0135124.
99US-0135353.
99US-0135629.
99US-0136021.
99US-0136392.
99US-0136782.
99US-0137222.
99US-0137528.
99US-0137502.
99US-0137724.
99US-0138094.
99US-0138540.
99US-0138847.
99US-0139119.
99US-0139452.
99US-0139453.
99US-0139454.
99US-0139455.
99US-0139456.
99US-0139457.
99US-0139458.
99US-0139459.
99US-0139460.
99US-0139461.
99US-0139462.
99US-0139463.
99US-0139750.
99US-0139763.
99US-0139817.
99US-0139899.
99US-0140353.
99US-0140354.
99US-0140895.
99US-0140823.
99US-0140991.
99US-0141287.
99US-0141842.
99US-0142154.
99US-0142055.
99US-0142390.
99US-0142803.
99US-0142920.
99US-0142977.
99US-0143542.
99US-0143624.
99US-0144005.
99US-0144085.
99US-0144086.
99US-0144325.
99US-0144331.
99US-0144332.
99US-0144333.
99US-0144334.
99US-0144335.
99US-0144352.
99US-0144632.
```


Search completed: June 2, 2002, 18:05:55
Job time: 253 sec

XX 30-MAR-2001; 2001WO-US08631.
XX
XX 31-MAR-2000; 2000US-0540217.
PR 23-AUG-2000; 2000US-0649167.
XX
XX (HYSE-) HYSEQ INC.
XX
XX Drmanac RT, Liu C, Tang YT;
PI
XX
XX WPI; 2001-639362/73.
DR N-PSDE; AAS82457.
XX

XX New isolated polynucleotide and encoded polypeptides, useful in
PT diagnostics, forensics, gene mapping, identification of mutations
PT responsible for genetic disorders or other traits and to assess
PT biodiversity
XX

PS Claim 20; SEQ ID No 48629; 103pp; English.

XX The invention relates to isolated polynucleotide (I) and
XX polypeptide (II) sequences. (I) is useful as hybridisation probes,
CC polymerase chain reaction (PCR) primers, oligomers, and for chromosome
CC and gene mapping, and in recombinant production of (II). The
CC polynucleotides are also used in diagnostics as expressed sequence tags
CC for identifying expressed genes. (I) is useful in gene therapy techniques
CC to restore normal activity of (II) or to treat disease states involving
CC (II). (II) is useful for generating antibodies against it, detecting or
CC quantitating a polypeptide in tissue, as molecular weight markers and as
CC a food supplement. (II) and its binding partners are useful in medical
CC imaging of sites expressing (II). (I) and (II) are useful for treating
CC disorders involving aberrant protein expression or biological activity.
CC The polypeptide and polynucleotide sequences have applications in
CC diagnostics, forensics, gene mapping, identification of mutations
CC responsible for genetic disorders or other traits to assess biodiversity
CC and to produce other types of data and products dependent on DNA and
CC amino acid sequences. ABG0010-ABG30377 represent novel human
CC diagnostic amino acid sequences of the invention.

CC Note: The sequence data for this patent did not appear in the printed
CC specification, but was obtained in electronic format directly from WIPO
CC at ftp.wipo.int/pub/published_pct_sequences.

XX Sequence 936 AA:

Query Match 5.5%; Score 103.5; DB 22; Length 936;
Best Local Similarity 22.4%; Pred. No. 0.39;
Matches 70; Conservative 49; Mismatches 84; Indels 109; Gaps 19;
QY 127 ILKDNQYELFADGEAMLFSAHLNSAIKPKPLYPHLIKAQREK-----RFRYISYL 178
Db 183 lledngw---lad-ealiy---vesevengiptvpanwsihrekvaafagaepphyinh 235
QY 179 LGL-----RELKAIKLVFEGKVT---LKAV-----KDIEAVPV---W- 210
Db 236 lgfprvgrlrelkkaqdsyagnstreeallavgrrelrarhwdqkqagidlpyqdfaw 295
QY 211 -----VAVNTAVMLGTGRPLMNPKKVA---SWIEDKNILL 244
Db 296 dhvlttslllgnvparhqnkdgsvidtlfirgraptgepaataekw----- 346
QY 245 YGTDTIEFGYRDIAQ--YRMSVEGLLEVIDINSELCLPSELKHSRGELYLTSSWAPDK 302
Db 347 fctnyhymvpefvgkqgqfklwtqlldvdeala-----lghkvkpvilgpytw--- 395
QY 303 SLRIWR---EDEGNARLNMLSYNRMGELALLAENSARG--W-----LPERRLD 348
Db 396 ---lwlgkvkgqefdrslldilpvyqqlaiaela-krgiewvqidepalvlelpgpwl 451
QY 349 AFRAIYNDRGE 360
Db 452 aykpaydalqgg 463
